ТРУДЫ

ИНСТИТУТА СИСТЕМНОГО ПРОГРАММИРОВАНИЯ РАН

PROCEEDINGS OF THE INSTITUTE FOR SYSTEM PROGRAMMING OF THE RAS

ISSN Print 2079-8156
Том 36 Выпуск 6

ISSN Online 2220-6426
Volume 36 Issue 6

Институт системного
программирования
им. В.П. Иванникова РАН

Москва, 2024

ИСП РАН

ISSN 2079-8156 (Print), ISSN 2220-6426 (Online)

Труды Института системного программирования РАН
Proceedings of the Institute for System Programming of the RAS

Труды ИСП РАН – это издание с двойной анонимной системой рецензирования, публикующее научные статьи, относящиеся ко всем областям системного программирования, технологий программирования и вычислительной техники. Целью издания является формирование научно-информационной среды в этих областях путем публикации высококачественных статей в открытом доступе.
Издание предназначено для исследователей, студентов и аспирантов, а также практиков. Оно охватывает широкий спектр тем, включая, в частности, следующие:

- операционные системы;
- компиляторные технологии;
- базы данных и информационные системы;
- параллельные и распределенные системы;
- автоматизированная разработка программ;
- верификация, валидация и тестирование;
- статический и динамический анализ;
- защита и обеспечение безопасности ПО;
- компьютерные алгоритмы;
- искусственный интеллект.

Журнал издается по одному тому в год, шесть выпусков в каждом томе.
Поддерживается открытый доступ к содержанию издания, обеспечивая доступность результатов исследований для общественности и поддерживая глобальный обмен знаниями.
Труды ИСП РАН реферируются и/или индексируются в:

Proceedings of ISP RAS are a double-blind peer-reviewed journal publishing scientific articles in the areas of system programming, software engineering, and computer science. The journal's goal is to develop a respected network of knowledge in the mentioned above areas by publishing high quality articles on open access.
The journal is intended for researchers, students, and practitioners. It covers a wide variety of topics including (but not limited to):

- Operating Systems.
- Compiler Technology.
- Databases and Information Systems.
- Parallel and Distributed Systems.
- Software Engineering.
- Software Modeling and Design Tools.
- Verification, Validation, and Testing.
- Static and Dynamic Analysis.
- Software Safety and Security.
- Computer Algorithms.
- Artificial Intelligence.

The journal is published one volume per year, six issues in each volume.
Open access to the journal content allows to provide public access to the research results and to support global exchange of knowledge. **Proceedings of ISP RAS** is abstracted and/or indexed in:

УДК004.45

Адрес: 109004, г. Москва, ул. А. Солженицына, дом 25.
Телефон: +7(495) 912-44-25
E-mail: info-isp@ispras.ru
Сайт: http://www.ispras.ru/proceedings/

Address: 25, Alexander Solzhenitsyn st., Moscow, 109004, Russia.
Tel: +7(495) 912-44-25
E-mail: info-isp@ispras.ru
Web: http://www.ispras.ru/en/proceedings

Содержание

Table of Contents

DOI: 10.15514/ISPRAS-2024-36(6)-1

Improving Estimation Models by Merging Independent Data Sources

[1] *F. Valdés-Souto, ORCID: 0000-0001-6736-0666 <fvaldes@ciencias.unam.mx>*
[2] *J. Valeriano-Assem, ORCID: 0009-0008-6473-1271 <jorge.valeriano@spingere.com.mx>*
[1] *National Autonomous University of Mexico Science Faculty, CDMX, México.*
[2] *SPINGERE, CDMX, México.*

Abstract. Software cost/effort estimation has been a key research topic for over six decades due to its industry impact. Despite numerous models, regression-based approaches dominate the literature. Challenges include insufficient datasets with enough data points and arbitrary integration of different source databases. This study proposes using the Kruskal-Wallis test to validate the integration of distinct source databases, aiming to avoid mixing unrelated data, increase data points, and enhance estimation models. A case study was conducted with data from an international company's Mexico office, which provides software development for "Microservices and APIs." Data from 2020 were analyzed. The estimation model's quality improved significantly. MMRE decreased by 25.4% (from 78.6% to 53.2%), standard deviation dropped by 97.2% (from 149.7% to 52.5%), and the Pred (25%) indicator rose by 3.2 percentage points. The number of data points increased, and linear regression constraints were met. The Kruskal-Wallis test effectively improved the estimation models by validating database integration.

Ключевые слова: linear regression model; software estimation; effort estimation; cost estimation; functional size; COSMIC method; Kruskal-Wallis.

For citation: Valdés-Souto F., Valeriano-Assem J. Improving estimation models by merging independent data sources. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 7-18. DOI: 10.15514/ISPRAS-2024-36(6)- 1.

Совершенствование моделей оценки путем объединения независимых источников данных

[1] *Ф. Вальдес-Соуто, ORCID: 0000-0001-6736-0666 <fvaldes@ciencias.unam.mx>*
[2] *Х. Валериано-Ассем, ORCID: 0009-0008-6473-1271 <jorge.valeriano@spingere.com.mx>*
[1] *Национальный автономный университет Мексики, факультет Науки, Мехико, Мексика.*
[2] *SPINGERE, Мехико, Мексика.*

Аннотация. Оценка затрат/усилий на программное обеспечение является ключевой темой исследований более шести десятилетий из-за его влияния на отрасль. Несмотря на многочисленные модели, подходы, основанные на регрессии, доминируют в литературе. Проблемы включают в себя недостаточные наборы данных с достаточным количеством точек данных и произвольную интеграцию различных исходных баз данных. В этом исследовании предлагается использовать тест Крускала-Уоллиса для проверки интеграции отдельных исходных баз данных с целью избежать смешивания несвязанных данных, увеличения точек данных и улучшения моделей оценки. Было проведено тематическое исследование с данными из офиса международной компании в Мексике, который обеспечивает разработку программного обеспечения для «микросервисов и API». Были проанализированы данные за 2020 год. Качество модели оценки значительно улучшилось. MMRE снизился на 25,4% (с 78,6% до 53,2%), стандартное отклонение снизилось на 97,2% (с 149,7% до 52,5%), а показатель Pred (25%) вырос на 3,2 процентных пункта. Количество точек данных увеличилось, и были соблюдены ограничения линейной регрессии. Тест Крускала-Уоллиса эффективно улучшил модели оценки, подтвердив интеграцию базы данных.

Ключевые слова: модель линейной регрессии; оценка программного обеспечения; оценка затрат; оценка стоимости; функциональный размер; метод COSMIC; тест Крускала-Уоллиса.

Для цитирования: Вальдес-Соуто Ф., Валериано-Ассем Х. Совершенствование моделей оценки путем объединения независимых источников данных. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 7–18 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–1.

1. Introduction

Software cost and effort estimation is crucial for planning, budgeting, and project success in software engineering. Regression-based estimation approaches are common, but the literature highlights challenges such as small datasets and the arbitrary combination of different source databases without proper validation. These issues are prevalent in both academia and industry, where small datasets are more common than expected. Estimation has potential for significant contributions to software engineering, particularly through leveraging statistical methods. This paper proposes a method for using established statistical techniques to validate the integration of distinct databases, thereby improving estimation models by increasing the number of data points. The paper outlines background information, the proposed procedure for data integration and validation, and the improvements observed in the estimation model, concluding with a discussion of the findings [1-4].

2. Background

2.1 Parametric Software Estimation

Software estimation, which began in the 1950s, has been crucial to the success of development projects, influencing budgeting and planning. Over more than 70 years, various techniques and classifications of estimation methods have been developed. However, many challenges and unanswered questions remain in software estimation research. A key factor in estimation accuracy is the measurement of software size [5]. Today, functional size is the only software feature that can be consistently quantified, emphasizing its importance. Every estimation model is closely tied to the method used to measure the input variables that produce the estimate.

2.1.1 Database Conformation for Parametric Estimation

When creating an estimation model, it's crucial to integrate a reference database from past projects. This database helps identify correlations between variables, with functional size being primary. However, regression-based models often face replication issues, as noted by various authors [1-4, 6-10]. These models are typically based on previous projects, but if a new project differs significantly from those in the database, the model's forecasts may be inaccurate. Challenges in accessing and interpreting data, as well as the limitations of available datasets, further complicate the process.

One major problem is the insufficient number of data points in datasets, which is critical for statistical reliability. Carbonera et al. classified datasets based on the number of points: high quality (more than 15), medium quality (10-15), and low quality (fewer than 10). The central limit theorem suggests that at least 30 data points are needed for each variable to approximate a normal distribution effectively.

Collecting 30 similar projects is often difficult, as noted by Morgenshtern et al., who highlight the cost and time involved in gathering historical data. Furthermore, combining data from different sources without proper evaluation can compromise its utility.

Organizations like the International Software Benchmarking Standards Group (ISBSG) and the Mexican Software Metrics Association (AMMS) maintain databases of past projects. The AMMS dataset, which includes data from real Mexican industry projects, shares similar features with the ISBSG dataset. Addressing the issue of limited data points may require developing techniques to integrate distinct databases using statistical methods, thereby improving the reliability of estimation models.

2.1.2 Estimation Models Performance Comparison

The performance of estimation models is assessed by applying quality criteria to evaluate their accuracy. Discrepancies between estimated and actual values are measured using criteria like Mean Magnitude of Relative Error (MMRE), Standard Deviation of MRE (SDMRE), Prediction level (PRED), Median Magnitude of Relative Error (MdMRE), and Mean Absolute Residual (MAR). Researchers have analyzed these techniques and identified various concerns regarding their effectiveness and reliability [10].

2.2 Kruskal-Wallis test

The Kruskal-Wallis [11, 12, 13] test is a nonparametric method used to compare the distributions of independent groups, serving as an alternative to one-way ANOVA when assumptions of normality and homogeneity are violated or when data are ordinal. Introduced by William Kruskal and Wilson Wallis in 1952, it ranks data from all groups and calculates a test statistic H based on these ranks. A higher H value suggests more evidence against the null hypothesis of no difference among distributions. H follows a chi-square distribution with $k-1$ degrees of freedom under the null hypothesis, where k is the number of groups. The test assesses whether the sample rank distributions differ significantly, indicating differences in population medians. If H exceeds the critical chi-square value, it implies significant differences among groups, leading to rejection of the null hypothesis. It is widely used in experimental and observational studies.

2.3 Outliers identification using Tukey test

The Tukey [12, 13] test, developed by John Tukey, is used to identify outliers in a dataset. It involves calculating the interquartile range (IQR) by subtracting the first quartile (Q1) from the third quartile (Q3). A threshold, typically 1.5 times the IQR, is set to flag outliers. Observations falling below Q1 minus the threshold or above Q3 plus the threshold are considered potential outliers. The test is robust against moderate deviations from normality and is effective for skewed or non-normally distributed data.

3. Case study: integrating distinct sources databases

This section presents a summary of a case study carried out at an international company with a Mexico office, referred to as COMPANY for confidentiality purposes. The office offers software development services to a financial institution, with data gathered in 2020.

The case study comprises three steps, but this paper will focus only on the final two:

1. **Project Identification/Classification and Functional Size Approximation:** COMPANY carried out this step to determine the types of projects that required estimation. They selected projects from a technological tower labeled "Microservices and APIs". Using the integrated information, we applied the EPCU approximation method [14] to measure the FURs of each project using COSMIC (ISO/IEC 19761).
2. **Incorporation of Additional Projects from Other Sources:** Since the COMPANY's provided projects were insufficient to develop a reliable estimation model, we sought out similar projects related to Microservices or API development in the ISBSG and AMMS databases. A total of forty-nine (49) projects were identified: 15 from the ISBSG database and 34 from the AMMS database.
3. **Constructing the Final Estimation Model:** To develop the final estimation model, we utilized the Kruskal-Wallis test to compare the distributions of independent groups and assess the feasibility of integration. We then followed the steps outlined by Valdés-Souto et al. in [7-9] to build and refine the estimation model.

The COMPANY and alternative sources (ISBSG, AMMS) are crucial in our project characterization process. They provide the essential data needed to select projects with similar characteristics, allowing us to compare size and effort. The effort was measured using COSMIC (ISO/IEC 19761), which serves as the fundamental metric for our project characterization. In Table 1.a), column 1 lists the acronym of the source from which the projects were obtained, column 2 shows the number of projects in the sample, and column 3 indicates the proportion of each group relative to the total number of projects. All the projects involved microservices or API development. Table 1.b) presents in column 1 the source acronym, in column 2 the functional size in CFP per group, and in column 3 the proportion of size per group relative to the total functional size in the sample.

Table 1. a) Sample size by source, "Microservices and APIS" projects. b) Total functional size by source.

SOURCE	Sample Size	%
COMPANY	8	14.0%
ISBSG	15	26.3%
AMMS	34	59.7%
Total	**57**	**100.0%**

a)

SOURCE	COSMIC Functional Size (CFP)	%
COMPANY	2418.7	11.0%
ISBSG	3873	17.6%
AMMS	15674.6	71.4%
TOTAL	**21966.3**	**100.0%**

b)

Based on the tables above, it can be seen that the AMMS database is a major contributor, representing 71.4% of the total functional size and 59.7% of the total projects. The ISBSG database is the second largest contributor in terms of project quantity and size, accounting for 26.3% of the total projects and 17.4% of the total size. Data from the COMPANY had the smallest contribution in both size and quantity. Due to the central limit theorem, the number of projects from the

COMPANY is insufficient to create a significant estimation model using only the initially provided data from the COMPANY.

Originally, the estimation model that could be developed using only the COMPANY's data is shown in Fig. 1, where the x-axis represents CFP and the y-axis represents effort. Although this model achieved an R^2 of over 77%, the limited amount of data prevents meaningful extrapolation of the conclusions.

The model with three datasets is shown in Fig. 2. In this case the model presents a R2 of 62% that it is lower than 77% of the initial model. The main concern is whether the additional dataset has a different distribution or if its mean significantly varies from the mean of the previous data. If so, the impact could be significant, and it might not be a good idea to integrate the data.

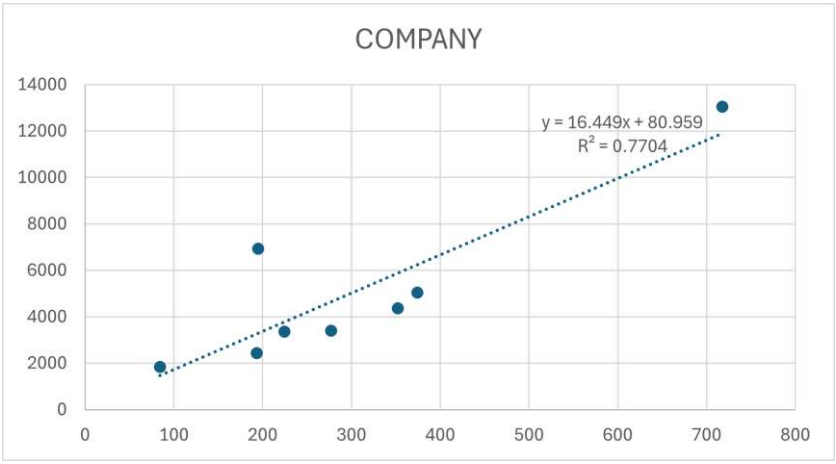


Fig. 1. COMPANY estimation model.

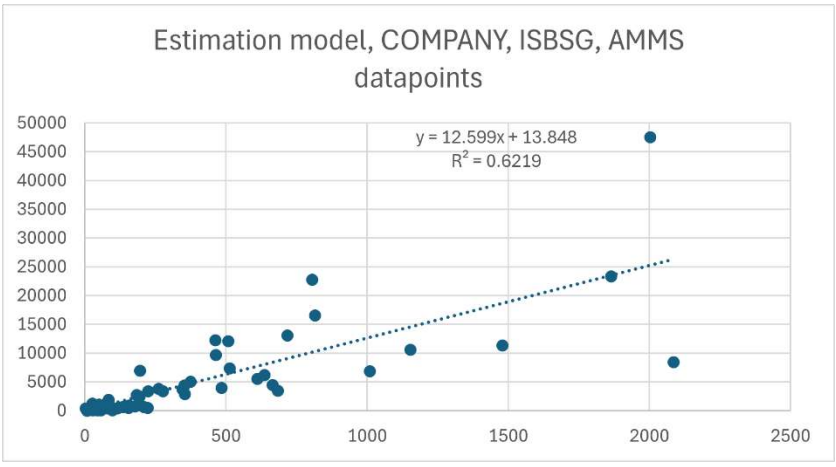


Fig. 2. COMPANY, ISBSG, AMMS estimation model.

The next step was to compare the distributions of independent groups to evaluate whether the integration was feasible with solid statistical foundations. Specifically, we assessed whether the distributions of the three databases (COMPANY, ISBSG, AMMS) for the PDR variable (HH/CFP) are the same or different. This evaluation helps determine whether it is appropriate to combine the three databases into a single database and build estimation models. Since the project samples come from different databases, they are considered independent samples in statistics. In this case, there are three samples. To assess these, we used a nonparametric test called the Kruskal-Wallis test [11],

[13], which allows us to conclude whether the distributions of the three samples are equal or different. The null hypothesis (H0) is that there is no significant difference between the distributions of the COMPANY, ISBSG, and AMMS databases. The alternative hypothesis (H1) is that at least one of the distributions from the COMPANY, ISBSG, or AMMS databases is significantly different. The significance level required is α=0.05. If the test's p-value is greater than or equal to 0.05, then H0 is correct; otherwise, if it is less than 0.05, H1 is correct. The Kruskal-Wallis test was performed using SPSS® version 25 in Spanish.

Table 2 summarizes the results of the Kruskal-Wallis test for the COMPANY, ISBSG, and AMMS databases. The p-value is less than 0.05 (LINE 3); therefore, the null hypothesis (H0) is rejected. Consequently, we conclude that there is a significant difference in at least one of the distributions of the COMPANY, ISBSG, and AMMS databases, as stated by the alternative hypothesis (H1). To determine which databases have different distributions, it is necessary to perform pairwise comparisons using the Kruskal-Wallis test, adjusting the resulting p-values to account for the number of tests. This adjustment is known as the Bonferroni correction [13]. Table 3 displays the results for each pair of datasets analyzed. The AMMS – COMPANY pair is the only one with an adjusted p-value (0.6171) greater than 0.01667 (0.05/3). From this, we conclude that the AMMS and COMPANY databases have the same distribution, while the ISBSG database has a different distribution. Therefore, it is only possible to integrate the COMPANY and AMMS datasets to build the estimation model, resulting in a total of 42 datapoints (COMPANY (8), AMMS (34)).

Table 2. Kruskal-Wallis test results for three datasets.

N	57
Degrees of freedom (Number of sets -1)	2
Asymp.sig. (p-value)	0.00001696

Table 3. Kruskal-Wallis test results by couple of datasets.

Pair	Asymp.sig. (p-value)	Asymp.sig. (p-value) with Bonferroni correction
ISBSG – AMMS	0.00001578	0.00004735
ISBSG – COMPANY	0.0006197	0.001859
AMMS – COMPANY	0.2057	0.6171

Once the integration validation is performed, we have the final dataset (COMPANY + AMMS) to develop an estimation model directly. The results are shown in Fig. 3. The generated estimation model is **$y = 8.6672x + 1586.9$**, with a determination coefficient **$R^2 = 0.5388$**. However, it is necessary to develop a linear regression model validation and diagnostics. The Normal probability graph and the Residuals graph were obtained using an Excel add-in to analyze the regression model, as shown in Fig. 4.

Fig. 4 shows evidence against normality, as the points do not follow the identity line in the normal probability graph. Additionally, in the residuals graph, the variance of the residuals increases with the fitted values, showing a systematic pattern and indicating non-constant variance, which means the data do not exhibit homoscedasticity. To correct the model's assumptions, we applied a logarithmic transformation to the functional size and effort variables and built a new estimation model. Refer to Fig. 5, where the x-axis represents Log(CFP) and the y-axis represents Log(effort).

The new estimation model is $Log(y) = 0.9326\ Log(x) + 2.8916$, with a coefficient of determination $R^2 = 0.8339$. We conducted validation and diagnostics using the transformed data, with results

shown in Fig. 6. The plot of fitted values against residuals indicates constant variance, as the dots do not display patterns, demonstrating homoscedasticity. Consequently, the estimation model in Fig. 5 meets the statistical principles of normality and homoscedasticity, making the linear regression model appropriate for this dataset.

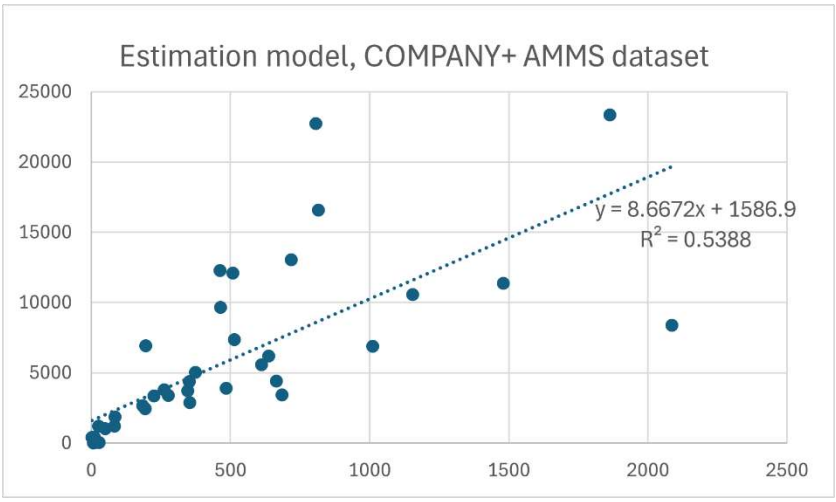


Fig. 3. Initial estimation model AMMS-COMPANY dataset.

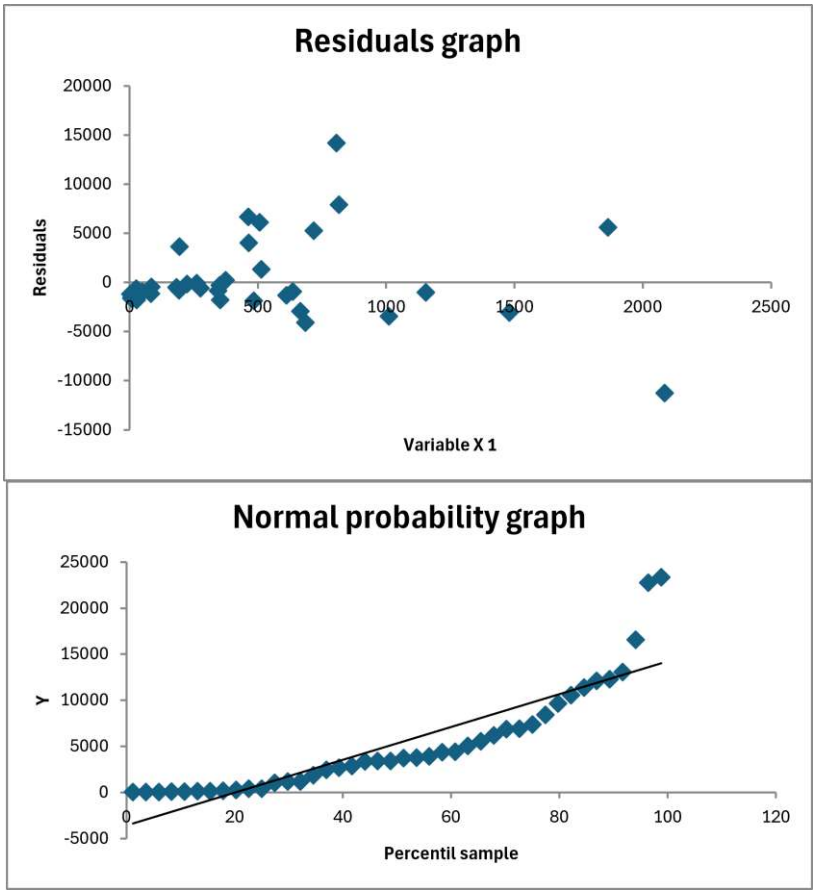


Fig. 4. Graph for validation and diagnostics AMMS-COMPANY dataset.

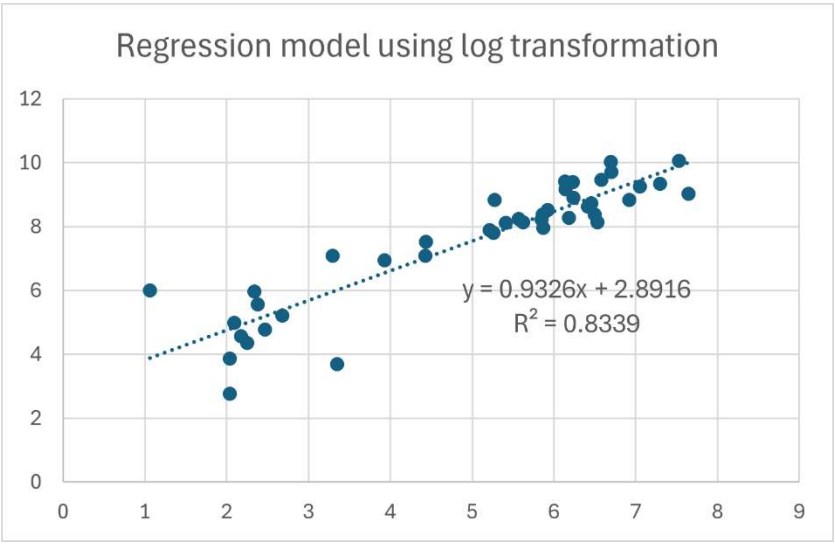


Fig. 5. Estimation model AMMS-COMPANY dataset using logarithm transformation.

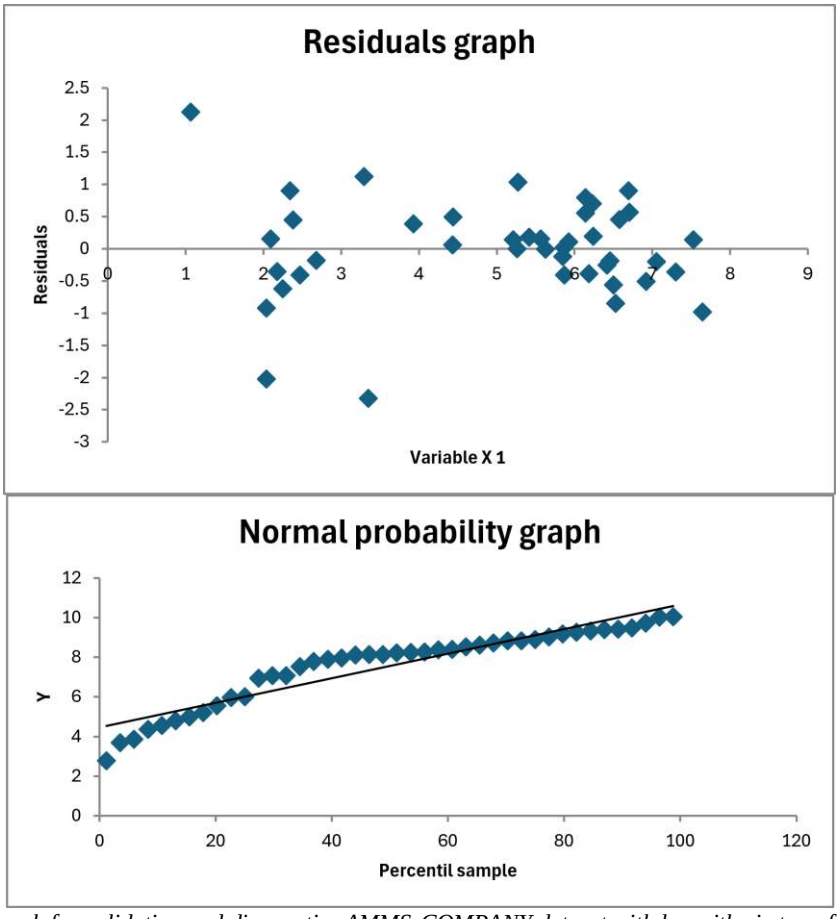


Fig. 6. Graph for validation and diagnostics AMMS-COMPANY dataset with logarithmic transformation.

After that, we search for outliers using the Tukey test, finding four (4) outliers as shown in Fig. 7. After removing the outliers, a new estimation model was obtained, the result is:

$$\text{Log}(y) = 0.9377\ \text{Log}(x) + 2.8996$$

with a Determination coefficient $R^2 = 0.9023$. The model uses logarithmic variables. To apply it to the actual variables, we need to eliminate the logarithmic transformation using the inverse operation (Euler's number, *e*), resulting in the final model: $\mathbf{y = x^{0.\,9377} * e^{2.8996}}$.

Table 4 presents the quality criteria for the developed estimation models. The best model is the last one, achieved after applying validations and diagnostics, then performing a transformation and removing the outliers.

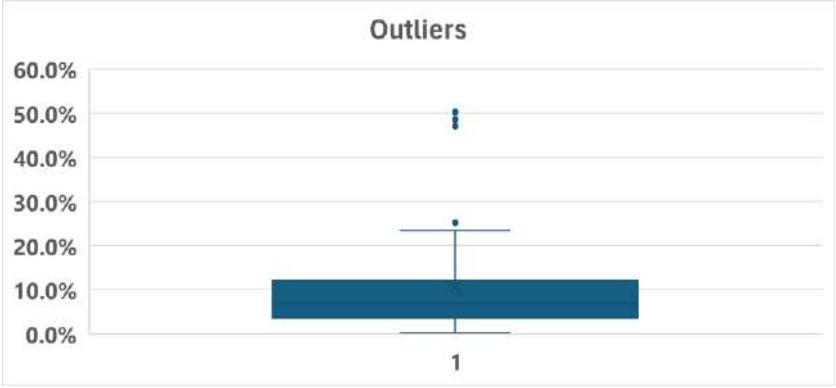


Fig. 7. Outliers AMMS-COMPANY dataset with logarithmic transformation.

Table 4. Quality criteria comparison for estimation models.

	$y = 16.449x + 80.959$ $R^2 = 0.7704$	$y = 8.6672x + 1586.9$ $R^2 = 0.5388$	$y = x^{0.9422} * e^{3.0156}$
N	8	42	38
MMRE	27.8%	78.6%	53.2%
SDMRE	13.4%	149.7%	52.5%
Pred (25%)	50.0%	31.0%	34.2%
Enough data	NO	YES	YES

4. Analysis

In the case study presented, the COMPANY under study had only eight (8) data points. Two additional datasets were considered: the ISBSG dataset with fifteen (15) data points and the AMMS dataset with thirty-four (34) data points.

However, the ISBSG data points were rejected by the Kruskal-Wallis test, resulting in a final dataset with forty-two (42) data points. After removing outliers, the dataset contained thirty-eight (38) data points.

The results obtained are related to MMRE that was reduced by 25.4% (from 78.6% to 53.2%), the standard deviation was reduced by 97.2% (from 149.7% to 52.5%), and the Pred (25%) indicator increased by 3.2 percentage points. Notably, the number of data points was significantly increased, from 8 to 38 (475%), bolstering the robustness of our findings.

5. Conclusions

Software cost/effort estimation has been a key research issue for over 60 years, with regression-based methods being widely used. However, issues have arisen related to dataset conformity,

including insufficient data points and arbitrary combining of different sources. This study presents a real case applying statistical methods, specifically the Kruskal-Wallis test, to determine if data from different sources can be integrated without compromising dataset integrity. The integration, validated through this analysis, allows for a larger and more significant dataset, improving the estimation models.

The case study demonstrated that an estimation model generated from validated integrated datasets outperformed one created from unvalidated sources. This underscores the importance of validation and diagnostic analysis in integration efforts. The study aimed to establish a formal methodology for creating reliable estimation models from diverse data sources, addressing a common issue in both industry and academia. While the proposed approach showed promise, it is crucial to apply statistical principles correctly; otherwise, the models might be ineffective. The methodology, developed and applied in the study, represents a significant advance in addressing the problem of small dataset sizes in software estimation.

References

[1]. M. Jørgensen and M. Shepperd, "A systematic review of software development cost estimation studies," IEEE Trans. Softw. Eng., vol. 33, no. 1, pp. 33–53, 2007, doi: 10.1109/TSE.2007.256943.

[2]. P. L. Braga, A. L. I. Oliveira, and S. R. L. Meira, "Software Effort Estimation using Machine Learning Techniques with Robust Confidence Intervals," in 7th International Conference on Hybrid Intelligent Systems, 2007, no. October 2007. doi: 10.1109/his.2007.56.

[3]. C. E. Carbonera, K. Farias, and V. Bischoff, "Software development effort estimation: A systematic mapping study," IET Res. Journals, vol. 14, no. 4, pp. 1–14, 2020, doi: 10.1049/iet-sen.2018.5334.

[4]. A. Abran, Software Project Estimation: The Fundamentals for Providing High Quality Information to Decision Makers, 1st ed. John Wiley & Sons, 2015.

[5]. A. Abran, Software Metrics and Software Metrology. Hoboken, New Jersey: John Wiley & Sons, 2010.

[6]. N. Kinoshita, A. Monden, M. Tshunoda, and Z. Yucel, "Predictability classification for software effort estimation," in Proceedings - 2018 IEEE/ACIS 3rd International Conference on Big Data, Cloud Computing, Data Science and Engineering, BCD 2018, 2018, no. 1, pp. 43–48. doi: 10.1109/BCD2018.2018.00015.

[7]. F. Valdés-Souto, "Validation of supplier estimates using cosmic method," CEURInternational Work. Softw. Meas. Int. Conf. Softw. Process Prod. Meas. (IWSM Mensura 2019), vol. 2476, pp. 15–30, 2019.

[8]. F. Valdés-Souto and L. Naranjo-Albarrán, "Improving the Software Estimation Models Based on Functional Size through Validation of the Assumptions behind the Linear Regression and the Use of the Confidence Intervals When the Reference Database Presents a Wedge-Shape Form," Program. Comput. Softw., vol. 47, no. 8, pp. 673–693, 2021, doi: 10.1134/S0361768821080259.

[9]. F. Valdés-Souto, "Creating an Estimation Model from Functional Size Approximation Using the EPCU Approximation Approach for COSMIC (ISO 19761)," in Software Engineering: Methods, Modeling and Teaching, Volume 4, Editorial., C. Mario, Z. Jaramillo, C. Elena, D. Vanegas, and W. P. Charry, Eds. Bogotá, Colombia, 2017, p. 468.

[10]. L. Lavazza, "Accuracy Evaluation of Model-based COSMIC Functional Size Estimation," in ICSEA 2017: The Twelfth International Conference on Software Engineering Advances, 2017, no. c, pp. 67–72.

[11]. W. A. Kruskal, W. H., & Wallis, "Use of Ranks in One-Criterion Variance Analysis," J. Am. Stat. Assoc., vol. 47, no. 260, pp. 583–621, 1952, doi: https://doi.org/10.1080/01621459.1952.10483441.

[12]. J. W. T. W. Tukey, Exploratory Data Analysis, 1st ed. Addison & Wesley, 1977.

[13]. R. R. Wilcox, Introduction to Robust Estimation and Hypothesis Testing, 4th Editio. Academic Press, 2016. [Online]. Available: https://shop.elsevier.com/books/introduction-to-robust-estimation-and-hypothesis-testing/wilcox/978-0-12-804733-0

[14]. A. Abran et al., "Early Software Sizing with COSMIC: Experts Guide," vol. 2020, no. May. Common Software Measurement International Consortium (COSMIC), pp. 1–67, 2020. doi: 10.13140/RG.2.1.4195.0567.

Информация об авторах / Information about authors

Франсиско ВАЛЬДЕС-СОУТО имеет степень PhD в области инженерии программного обеспечения по специальности “Измерение и оценка программного обеспечения” в Высшей технологической школе (ETS) в Канаде, две магистерские степени в Мексике и Франции. Президент COSMIC. Доцент факультета наук Национального автономного университета Мексики. Основатель Мексиканской ассоциации метрик программного обеспечения (AMMS). Имеет более 25 лет опыта в разработке критически важного программного обеспечения. К настоящему времени опубликовал более 50 научных работ, включая статьи в индексированных журналах, трудах научных конференций, книгах и главах книг. Является главным промоутером проекта изучения формальных метрик программного обеспечения в Мексике, продвигая COSMIC (ISO/IEC 19761) в качестве национального стандарта. Член Национальной системы исследователей (SNI). Область научных интересов: измерение и оценка программного обеспечения, применяемого для управления проектами программного обеспечения, управление тематикой, производительностью и экономикой разработок программного обеспечения.

Francisco VALDÉS-SOUTO has a PhD degree in Software Engineering with a specialty in Software Measurement and Estimation at the École de Technologie Supérieure (ETS) in Canada, two master's degrees in Mexico and France. President of COSMIC. Associate Professor of the Faculty of Sciences of the National Autonomous University of Mexico (UNAM). Founder of the Mexican Association of Software Metrics (AMMS). More than 25 years of experience in critical software development. He currently has more than 50 publications including articles in Indexed Journals, Proceedings, books and book chapters. He is the main promoter of the topic of formal software metrics in Mexico, promoting COSMIC (ISO/IEC 19761) as a National Standard. Member of the National System of Researchers (SNI). Research interests: software measurement and estimation applied to software project management, scope management, productivity and economics in software projects.

Хорхе ВАЛЕРИАНО-АССЕМ – Магистр компьютерных наук и инженерии в Национальном автономном университете Мексики, специалист-консультант по формальному измерению и оценке программного обеспечения с 2016 года. Сфера научных интересов: метрики программного обеспечения (COSMIC), модели оценки программного обеспечения, модели валидации программного обеспечения, оценка функциональных и нефункциональных требований, оценка эффективности проектов разработки программного обеспечения на основе метрик программного обеспечения, оценка качества программных продуктов.

Jorge VALERIANO-ASSEM – master in Computer Science and Engineering from the National Autonomous University of Mexico, specialist consultant in formal software measurement and estimation since 2016. Areas of interest: Software metrics (COSMIC), Software estimation models, Software Validation Models, Estimation of Functional and Non-Functional Requirements, Evaluation of the Performance of Software Development Projects aligned to Software Metrics, Evaluation of the Quality of the Software Development Product.

DOI: 10.15514/ISPRAS-2024-36(6)-2

Class Balancing Approaches to Improve for Software Defect Prediction Estimations

Á. J. Sánchez-García, ORCID: 0000-0002-2917-2960 <angesanchez@uv.mx>
X. Limón, ORCID: 0000-0003-4654-636X <hlimon@uv.mx>
S. Domínguez-Isidro, ORCID: 0000-0002-9546-8233 <sauldominguez@uv.mx>
D. J. Olvera-Villeda, ORCID: 0009-0000-5615-1498 <olveradan@hotmail.com>
J. C. Pérez-Arriaga, ORCID: 0000-0003-2354-2462 <juaperez@uv.mx>

Faculty of Statistics and Informatics, Veracruz University,
Veracruz, Mexico.

Abstract. Addressing software defects is an ongoing challenge in software development, and effectively managing and resolving defects is vital for ensuring software reliability, which is in turn a crucial quality attribute of any software system. Software defect prediction supported by Machine Learning (ML) methods offers a promising approach to address the problem of software defects. However, one common challenge in ML-based software defect prediction is the issue of data imbalance. In this paper, we present an empirical study aimed at assessing the impact of various class balancing methods on the issue of class imbalance in software defect prediction. We conducted a set of experiments that involved nine distinct class balancing methods across seven different classifiers. We used datasets from the PROMISE repository, provided by the NASA software project. We also employed various metrics including AUC, Accuracy, Precision, Recall, and the F1 measure to gauge the effectiveness of the different class balancing methods. Furthermore, we applied hypothesis testing to determine any significant differences in metric results between datasets with balanced and unbalanced classes. Based on our findings, we conclude that balancing the classes in software defect prediction yields significant improvements in overall performance. Therefore, we strongly advocate for the inclusion of class balancing as a pre-processing step in this domain.

Keywords: software defect prediction; statistical analysis; imbalanced class; PROMISE; datasets; metrics; oversampling; undersampling.

For citation: Sánchez-García Á. J., Limón X., Domínguez-Isidro S., Olvera-Villeda D. J., Pérez-Arriaga J. C. Class balancing approaches to improve for software defect prediction estimations. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 19-38. DOI: 10.15514/ISPRAS-2024-36(6)-2.

Подходы к балансировке классов для улучшения оценок прогнозирования дефектов программного обеспечения

А. Х. Санчес-Гарсия, ORCID: 0000-0002-2917-2960 <angesanchez@uv.mx>
К. Лимон, ORCID: 0000-0003-4654-636X <hlimon@uv.mx>
С. Домингес-Исидро, ORCID: 0000-0002-9546-8233 <sauldominguez@uv.mx>
Д. Х. Олвера-Вийеда, ORCID: 0009-0000-5615-1498 <olveradan@hotmail.com>
Х. К. Перес-Арриага, ORCID: 0000-0003-2354-2462 <juaperez@uv.mx>

Университет Веракруса, факультет статистики и информатики, Веракрус, Мексика.

Аннотация. Постоянной проблемой в разработке программного обеспечения является устранение дефектов в этом обеспечении. И эффективное управление и устранение дефектов имеют жизненно важное значение для обеспечения надежности программного обеспечения, что, в свою очередь, является важнейшим атрибутом качества любой системы программного обеспечения. Прогнозирование программных дефектов, поддерживаемое методами машинного обучения (ML) – это многообещающий подход к решению проблемы программных дефектов. Тем не менее, одной из общих проблем в прогнозировании дефектов программного обеспечения на основе ML является проблема дисбаланса данных. В этой статье мы представляем эмпирическое исследование, направленное на оценку влияния различных методов балансировки классов на проблему дисбаланса классов в прогнозировании дефектов программного обеспечения. Мы провели ряд экспериментов, которые включали девять различных методов балансировки классов по семи различным классификаторам. Мы использовали наборы данных из репозитория PROMISE, предоставленные программным проектом NASA. Мы также использовали различные метрики, включая AUC, точность, полнота, отзыв и меру F1, чтобы оценить эффективность методов балансировки различных классов. Кроме того, мы применили проверку гипотез, чтобы определить любые существенные различия в метрических результатах между наборами данных со сбалансированными и несбалансированными классами. Основываясь на наших выводах, мы пришли к выводу, что балансировка классов в прогнозировании дефектов программного обеспечения дает значительное улучшение общей производительности. Поэтому мы решительно выступаем за включение балансировки классов в качестве этапа предварительной обработки в этой области.

Ключевые слова: прогнозирование дефектов программного обеспечения; статистический анализ; несбалансированный класс; репозиторий PROMISE; наборы данных; метрики; избыточная выборка; недостаточная выборка.

Для цитирования: Санчес-Гарсия А. Х., Лимон К., Домингес-Исидро С., Олвера-Вийеда Д. Х., Перес-Арриага Х. К. Подходы к балансировке классов для улучшения оценок прогнозирования дефектов программного обеспечения. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 19–38 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–2.

1. Introduction

This paper is an extension of work originally presented in 2023 11th International Conference in Software Engineering Research and Innovation (CONISOFT) [1]. The original work addressed a systematic literature review on class balancing methods in software defect prediction data sets.

Software Engineering is a discipline that aims to develop and deliver quality software products, through the execution of quality processes. Quality can be defined as “the degree to which a component, system or process satisfies the specified requirements and/or the needs and expectations of the user/customer” [2].

One of the most important attributes to satisfy customer requirements is reliability, which is defined as “the probability of failure- free operation in a specific environment during a given period of time” [3]. System defects, which are imperfections or deficiencies in a system or component that could prevent it from carrying out its intended purpose, can have an impact on this reliability [4].

This is important because “quality is inversely proportional to the number of defects found in a system” [5], including an impact on software reliability.

An area of software engineering that is gaining importance due to the need to increase the reliability of a software system is the well-known software defect prediction (SDP). This area aims to predict the presence of defects in the software during or after its development. To make these predictions, it is necessary to have historical data that allows generating prediction models based on some metrics taken from software projects.

One of the most used historical data repositories for defect prediction is the PROMISE Software Engineering repository [6], which consists of several datasets for different purposes, such as effort or defect prediction.

The main data sets for SDP in PROMISE repository are provided by NASA software projects, which are:

1) CM1, which is a NASA spacecraft instrument written in the C programming language.
2) JM1 is written in the C programming language, and it is from a real-time predictive ground system.
3) KC1 written in C++ programming language, is a system implementing storage management for receiving and processing ground data.
4) KC2 data are from C++ functions as part of the same project of KC1.
5) PC1 data from C functions for a flight software for earth orbiting satellite.

These five datasets include data obtained from McCabe [7] and Halstead [8] static code metrics. These data sets contain a class label that states whether the record is defect-prone or not, making it a classification problem.

However, as Table 1 shows, the number of records of each class (yes or no) is unbalanced, which impacts the precision of prediction models, since models bias toward the most common value, causing prediction errors, which can impact software engineering practitioners.

Table 1. Distribution of classes by data set.

Data set	Instances	Class Yes	Class No
CM1	498	49 (9.83 %)	449 (90.16%)
JM1	10,885	8,779 (80.65%)	2,106 (19.35%)
KC1	2,109	326 (15.45%)	1,783 (84.54%)
KC2	522	105 (20.5%)	415 (79.5%)
PC1	1,1109	1,032 (94.05%)	77 (6.94%)

It is important to highlight that class unbalance is a prevalent challenge in software defect prediction. The five datasets discussed in this paper serve as illustrative examples to emphasize the prominence of this issue.

When a data set is unbalanced, machine learning (ML) algorithms will more frequently predict values from the majority class to increase the value of predictions with the training data. This also implies that learning models could not generalize new data that is not part of the training data. In conclusion, these datasets serve as valuable reference points for researchers aiming to enhance classifiers for SDP, as well as for software engineers who seek optimal classifier options for this task. Nevertheless, it is crucial to recognize that the accuracy of these classifiers may be affected by bias arising from the number of examples in the majority class.

As stated, and demonstrated in [9], there is no ML algorithm that has the best performance for all data sets. Therefore, given the class unbalance nature of SDP, we propose to carry out an empirical and statistical study, applying different algorithms and approaches for class balancing to the data sets of SDP, with the aim of establishing which approaches can be used by software engineers to obtain better predictions of software defects, and consequently, deliver reliable and quality products.

Due to what was mentioned above and what is described in section 2 (Related work), our main contributions in this work are:

- Unlike similar studies, we experimented with three class balancing approaches: oversampling, undersampling, and a combination of them.
- We tested the class balancing algorithms with eight classification algorithms from different approaches such as decision trees, K-nearest neighbor, Bayesian approaches, neural networks and ensemble classifiers (both boosting and bagging).
- We provide details for our experiments, such as the versions of the classification algorithms, type of implementation and parameters used, so that the results can be replicated.
- The data sets used are public, unlike studies where their own data sets are used.
- Our results not only report the approach that showed the best performance during the experiments, but the results are statistically validated to know if in the prediction results, there is a significant difference in performance when balancing the data sets.

This paper is structured as follows: Section 2 mentions related work. Section 3 presents the background of class balancing algorithms. Section 4 describes the characteristics of the experimentation and evaluation. Section 5 analyzes the results obtained. Section 6 identifies some validity threats of this study. Finally, section 7 draws conclusions and presents future work.

2. Related work

In our literature review reported in the last 6 years (until 2024), 40 studies were found that implemented proposals to balance classes in data sets of software defect prediction, of which 60% used Oversampling techniques, 18% Undersampling, 15% Oversamplig + Undersampling, and 8% Ensemble approaches. Table 2 shows the studies grouped by their balancing approach.

Table 2. Distribution of studies according to the balancing class approach.

Approach	Studies
Oversampling	[10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38]
Undersampling	[39], [40], [41], [11], [17], [18], [20], [34], [36], [37]
Combination (oversampling + undersampling)	[42], [43], [44], [45], [46], [47]
Ensemble	[48], [49], [50]

Of the 40 studies identified, only three of them [34-36] carry out an experiment like our proposal.

Pandey and Tripathi in [34] presented an empirical study in which the influence, in terms of F-score and Mathews Correlation Coefficient (MCC), of four sampling techniques (Class Balancer, Resample technique, Synthetic minority oversampling technique (SMOTE), and Spread subsample) were evaluated in twelve different classifiers: Naïve Bayes (NB), Logistic Regression (LR), Multilayer Perceptron (MLP), Instance-based learning (IBK), AdaBoost, Bagging (Bagg), Logistic Boost (LB), Repeated incremental pruning to produce error reduction (RIPPER), Decision Tree (J48), and Random Forest (RF). The authors considered 22 datasets in the Software Defect Prediction domain in their experiments. In their findings, the authors highlighted that MCC is the most helpful performance metric for the imbalance dataset; the Resample technique gave the best results in most ML techniques. Finally, the authors leave open the investigation regarding the relationship between the classifier and the sampling methods.

On the other hand, Zhang et al. in [35] measured the effect of using four Oversampling techniques for class balancing in just-in-time SDP: i) Random Oversampling, ii) SMOTE, iii) Borderline-SMOTE, and iv) ADASYN. The authors proposed the OSNECL method, which combines Oversampling techniques with ensemble classification methods, such as Bagging, AdaBoost, Random Forest, and GBDT. In their experiments, six opensource projects written in C++ and Java

(Bugzilla, Columba, Eclipse JDT, Mozilla, Eclipse Platform, and Postgres) were used, from which 16 characteristics related to the defects were extracted with the SZZ algorithm. As a result of the experiments, it was found that the combination of Random Forest and Random Oversampling provided the best results.

However, the authors left it open to continue researching class balancing with Deep Ensemble learning methods to explore new ways to improve the balancing problem.

Finally, Yang et al. [36] conducted a study that evaluated the impact of sampling, Random Under/Oversampling, SMOTE, and OSS techniques for the class balancing problem on Deep learning-based vulnerability detection (DLVD). The DLVD techniques used in their experiments were Deving, Reveal, IVDetect, and LineVul. Additionally, they selected three benchmark datasets, with which different findings could be obtained, such as Oversampling outperforms Undersampling approaches, sampling on raw data outperforms sampling on feature space; generally, Random Oversampling on raw data performing the best among all studied sampling methods, including SMOTE and OSS.

Table 3 compares the sampling techniques analyzed in the studies where we identify them as similar to our experimentation, where it can be seen that the comparison was carried out with a maximum of four class balancing algorithms, none of them is a combination of oversampling and undersampling. Our study proposes the comparison of nine class balancing algorithms, including a combination of oversampling and undersampling.

Table 3. Sampling techniques analyzed in related works.

Sampling Method	**[34]**	**[35]**	**[36]**
Random undersampling			✓
Random Oversampling		✓	✓
SMOTE	✓	✓	✓
OSS			✓
Class balancer	✓		
Resample	✓		
Spread subsample	✓		
Borderline – SMOTE		✓	
ADASYN		✓	

Table 4 presents the learning techniques with which the techniques were combined. In the case of the study [36], they used specific deep- learning techniques to detect vulnerabilities, which is not directly related to SDP. We also include different classification approaches and algorithms.

Table 4. Learning methods studied in related works.

Learning classifier	**[34]**	**[35]**	**[36]**
Naïve Bayes (NB)	✓		
Multiplayer perceptron (MLP)	✓	✓	
K-Nearest Neighbor (KNN)	✓		
Decision Tree (DT)	✓		
AdaBoost (AB)	✓	✓	
Bagging	✓	✓	
Logistic Boost (LB)	✓		
RIPPER	✓		
Random Forest (RF)	✓	✓	
Gradient Boosting (GB)		✓	
Deep learning approaches			✓

3. Class balancing methods

We examined nine representative methods to address class imbalance in our experiments. These methods can be classified into three categories as it is shown in Fig. 1: oversampling, undersampling, and combined. Below, we provide an explanation of each method.

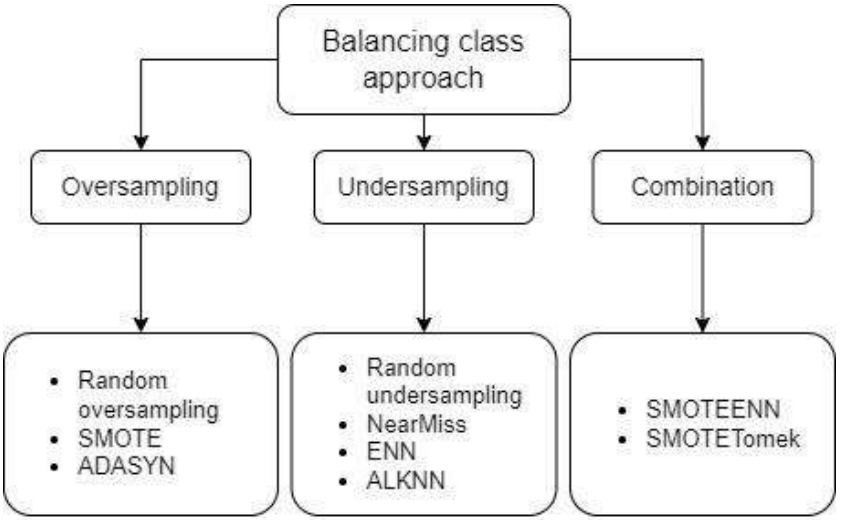


Fig. 1. Balancing class approaches chosen for experimentation.

3.1 Oversampling

Fig. 2 shows the strategy of class balancing by oversampling, where it consists of artificially increasing the number of samples of the minority class to balance the class distribution.

Three algorithms based on this approach are presented below.

1) Random over sampling: Also known as Random Over Sampling Examples (ROSE) [51], this simple method involves randomly selecting examples from the minority class or classes and repeating them with replacement. This process creates a new dataset that is balanced by increasing the representation of the minority class.
2) SMOTE [52], an acronym for Synthetic Minority Oversampling Technique, is a widely used oversampling method designed to address class imbalance. It involves the creation of synthetic data points in the feature space for the minority class. To over sample the minority class, SMOTE takes each minority class sample and generates synthetic examples along the line segments that connect the sample to its k-nearest neighbors from the same class. The value of k determines the number of nearest neighbors considered. The amount of oversampling required can be adjusted, and random selection is used to choose neighbors from the k-nearest neighbors.
3) ADASYN [53], short for Adaptive Synthetic Sampling Approach for Imbalanced Learning, is an oversampling technique specifically designed to address class imbalance. It involves the creation of synthetic data points for the minority class. The fundamental concept behind ADASYN is to use a weighted distribution to generate synthetic data based on the level of difficulty of learning for each minority class example. The algorithm identifies minority class examples that are harder to learn and generates a higher number of synthetic data points for them compared to examples that are easier to learn. ADASYN employs a density distribution based on the k-nearest neighbors of each minority class example to determine the number of synthetic samples to be generated.

3.2 Undersampling

Undersampling is another technique used to address class imbalance in data sets, where unlike oversampling, in this approach the number of samples of the majority class is reduced to equalize the class distribution with the minority class as can be seen in Fig. 3.

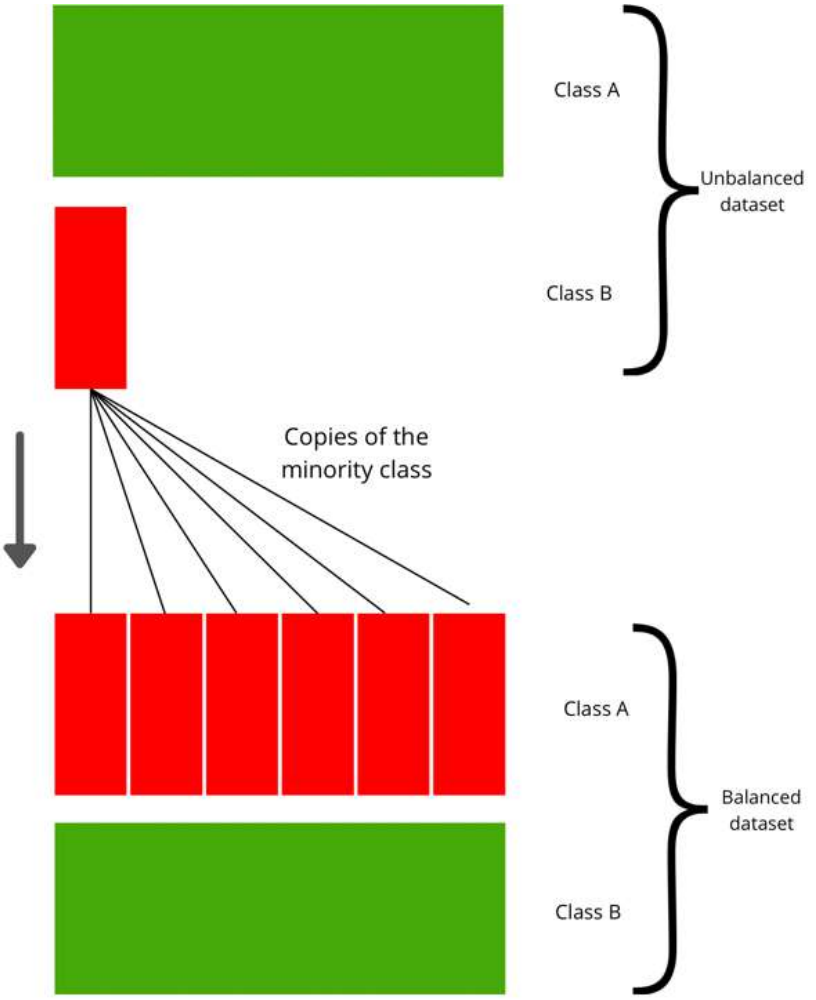


Fig. 2. Oversampling strategy.

Below, four selected undersampling methods are described.

1) Random undersampling is a technique used to reduce the number of majority class examples in a dataset. It is similar to random oversampling, but with the opposite objective. Random undersampling (RUS) [54] randomly selects examples from the majority class, either with or without replacement. The number of examples selected in random undersampling depends on the desired balancing goal. In a fully balanced dataset, the number of selected examples would equal the number of examples from the minority class.
2) NearMiss is an undersampling technique originally introduced in [55]. This method aims to balance class distribution by selecting a subset of majority class examples that are closest to the minority class examples based on their k-nearest neighbors. The NearMiss method

consists of three variants: NearMiss-1, NearMiss-2, and NearMiss-3.

- NearMiss-1: this variant focuses on selecting majority class examples that are closest to some of the minority class examples. It identifies majority class examples with the smallest average distance to their k-nearest minority class examples. The objective is to remove majority class examples that are farthest from the minority class.

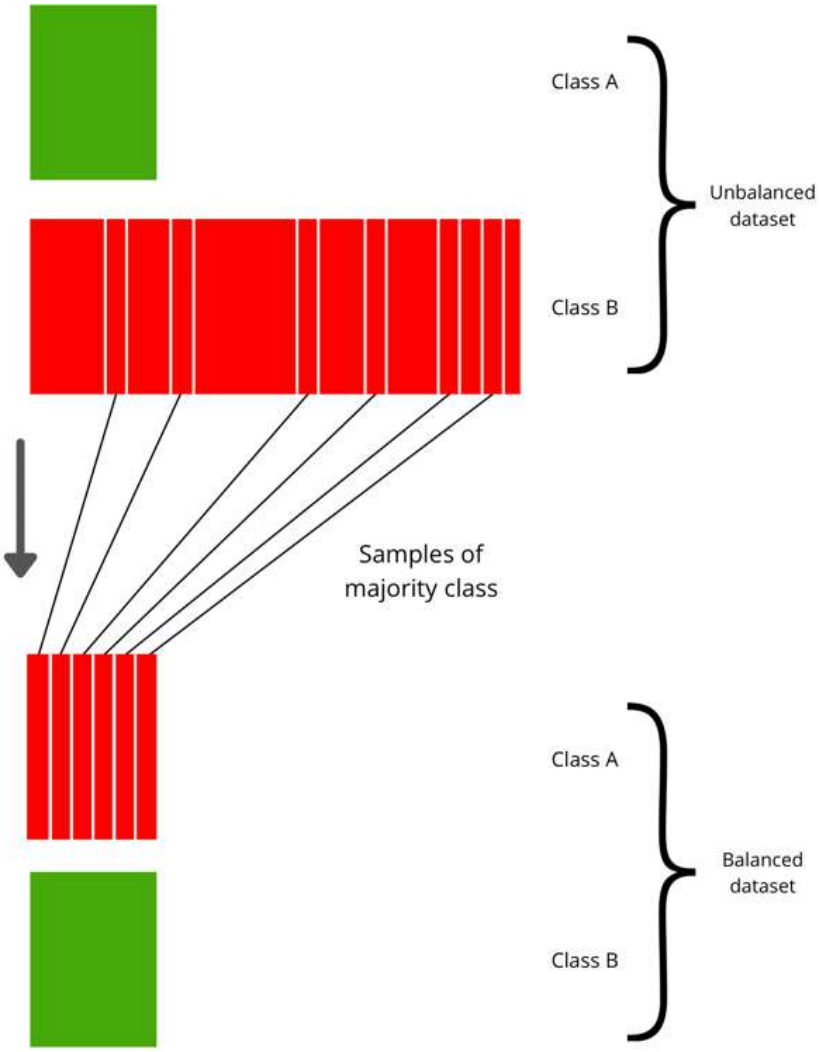


Fig. 3. Undersampling strategy.

- NearMiss-2: this variant selects majority class examples that are closest to all minority class examples. It chooses examples based on their average distances to the k farthest minority class examples.
- NearMiss-3: selects a specific number of closest majority class examples for each minority class example. The intention is to ensure that each minority class example is surrounded by some majority class examples.

3) ENN (Edited Nearest Neighbors), proposed by [56], is an undersampling method for

addressing class imbalance. The ENN algorithm operates by examining the k-nearest neighbors of each example in the dataset. If a majority class example is misclassified, it is removed from the dataset. On the other hand, if a minority class example is misclassified, its k neighbors belonging to the majority class are removed. The k value proposed in the original work is 3. It is important to note that ENN is primarily designed to remove noisy and ambiguous examples, rather than achieving perfect class balance.

4) AllKNN (All K-Nearest Neighbors), proposed in a study by [57], is an iterative undersampling method that extends the concept of ENN. Rather than using a fixed value of k, AllKNN repeatedly applies ENN with increasing values of k starting from 1. This iterative process continues until a maximum value of k is reached or until the minority class becomes the majority class. The use of progressively larger k values allows AllKNN to capture more information from the nearest neighbors and potentially achieve better imbalance reduction.

3.3 Combination

Finally, by combining oversampling and undersampling techniques, the strengths of each approach can be leveraged for better classification results. Two algorithms that combine the two previous approaches are described below.

1) SMOTEEN (SMOTE Edited Nearest Neighbors), proposed by [58], is a method that leverages both oversampling and undersampling techniques to address class imbalance. The SMOTEEN algorithm combines the synthetic minority oversampling technique (SMOTE) with the edited nearest neighbors (ENN) undersampling. It follows a two-step process to rebalance the dataset. First, SMOTE is applied to the minority class to generate synthetic instances, thereby increasing its representation. Next, ENN is employed to remove potentially noisy or mislabeled instances from both the minority and majority classes. The SMOTEEN process may be repeated until achieving a balanced dataset or reaching a stopping criterion.

2) SMOTETomek [59] is a combination of oversampling technique SMOTE, and undersampling technique Tomek Links [60]. SMOTETomek consists of two main steps. In the first step, SMOTE is applied to the minority class, oversampling with synthetic data and balancing the dataset. The second step of SMOTETomek utilizes Tomek Links. Tomek Links identify specific pairs of instances, one from the majority class and one from the minority class, that are nearest neighbors to each other. These pairs are ambiguous or potentially noisy instances, and thus are removed from the date set, undersampling at the same time examples from both classes.

4. Experimental design

Fig. 4 shows the steps followed for the experiments in this study. For our experiments, the only preprocessing performed was checking for missing values. In all five data sets, only JM1 included 5 records with at least one missing value. These records were removed.

After data preprocessing, the class balancing algorithms mentioned in Section 3 will be applied to each of the five data sets.

Subsequently, each classification learning algorithm described in section 4.1 will be tested on each data set (with and without class balancing).

The models are evaluated through a k-fold cross validation process in which the data set is divided into k sets (one subset is used for testing and the other k-1 subsets are used as training). This process is executed k times, such that each record is part of both a test and training set.

The metrics described in section 4.2 are applied to the models to evaluate the predictions. Finally, a statistical analysis is performed for each metric applied to each model developed by each classification algorithm and to each data set obtained with and without class balancing.

Finally, the results are statistically analyzed to empirically verify whether the results with the class balancing methods studied are significantly better.

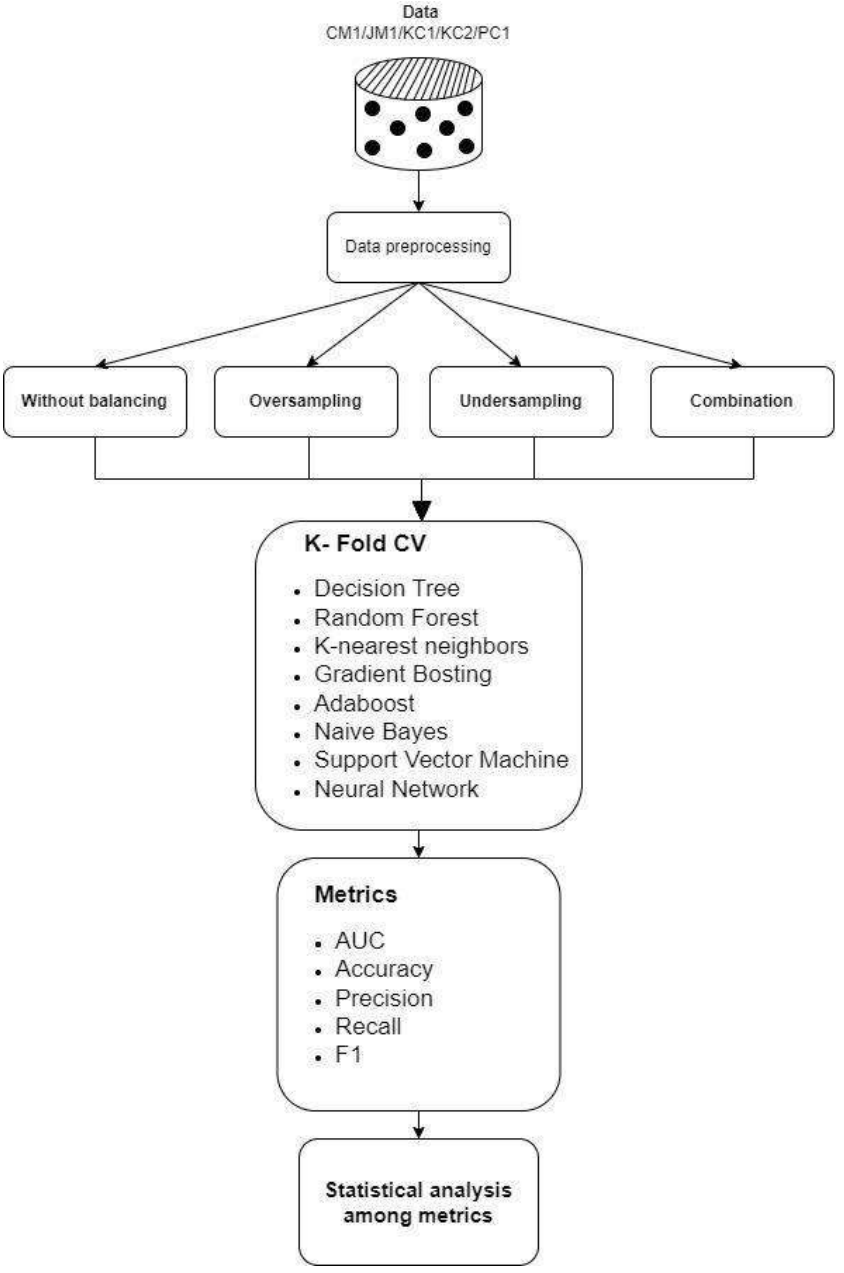


Fig. 4. Experimental design steps to validate models with balanced data sets.

4.1 Classification learning algorithms

As demonstrated in [9], there is no ML algorithm that performs better than the others for any data set. Therefore, we selected a variety of classification learning algorithms for experimentation, and they are described below, including their characteristics, advantages and disadvantages that could influence in the prediction results.

We chose several classification approaches, such as based on decision trees, class similarity (k-nearest neighbor), ensemble algorithms for both bagging and boosting, based on Bayesian approaches and finally on neural networks approach.

Decision Trees are a popular supervised learning method used for both classification and regression tasks [61]. They build a tree structure consisting of rules based on the input features, which guide the evaluation of new examples. This hierarchical representation enables visualizing the resulting model. Hence, Decision Trees employ a white model approach. Decision Trees have several advantages. Firstly, they require minimal data preparation, and they can handle both numeric and nominal data. Secondly, they are nonparametric, meaning they make no assumptions about the underlying data distribution. However, Decision Trees are prone to overfitting, to address this, pruning methods can be employed to simplify the tree and prevent overfitting. It is also important to note that Decision Trees can be affected by class imbalance [62].

K-Nearest Neighbors (KNN) [63] is a nonparametric and supervised method for classifying data instances. It utilizes a similarity metric, such as Euclidean distance, to identify the most similar examples and establish the concept of neighbors. By setting a parameter K, the algorithm determines the K-nearest neighbors. When classifying a new instance, a policy needs to be defined. A popular policy is majority vote, where the predicted class is the most common class among the K-nearest neighbors. It is important to note that the classifier's performance is sensitive to the choice of K. The KNN classifier may encounter limitations when dealing with large datasets. Additionally, it is important to consider feature scaling to ensure accurate results.

Random Forests [64] is an ensemble technique used for classification tasks. As an ensemble technique, it combines multiple classification models, specifically trees. Each tree is built from different random subsets of the dataset, allowing for improved generalization and robustness. One of the key features of Random Forests is the option to use a random sample of features when selecting the best splitting point in each tree. This further enhances the diversity and reduces correlation among the individual trees. During classification of a new example, each tree in the forest independently classifies the instance. Then, a decision policy, such as majority vote or averaging the predicted probabilities from each tree, is applied to determine the final class for the new instance. Random Forests are well-suited for handling large datasets with a high number of examples and features. They are also robust against data noise and overfitting.

AdaBoost [65] (Adaptive Boosting) is an iterative ensemble classification method that combines multiple weak classifiers to create a robust classifier. It utilizes a weighting strategy, where initially every example in the training set has equal weight. As the iterations progress, the weights of the examples are adjusted based on the classification performance of the classifiers. Additionally, AdaBoost assigns weights to each classifier based on their performance, with higher weights given to classifiers with better accuracy. During classification of a new instance, AdaBoost applies a weighted majority vote among the classifiers to determine the final class label.

Gradient Boosting [66] is an ensemble method used for both classification and regression tasks. It combines multiple weak models to create a stronger model. Gradient Boosting relies on Gradient optimization to iteratively build the ensemble by fitting weak models to the residuals. This process aims to correct errors in the current ensemble for Gradient optimization, residuals representing the gradients of the loss function are calculated with respect to the predicted probabilities. The subsequent weak model is then trained on these residuals to further improve the ensemble's performance. Additionally, each model in the ensemble is assigned a weight to assess its contribution. These weights consider the model's performance and are used to guide the ensemble

towards more accurate predictions. During the classification of a new example, a weighted vote policy is applied.

Naive Bayes: Naive Bayes is a supervised and probabilistic method used for classification tasks based on Bayes theorem [67]. It is a type of Bayesian Network method that operates under the assumption of feature independence. This strong hypothesis assumes that each feature in the data set is independent of each other given the class label. This assumption simplifies the model and allows for fast training as only the conditional probabilities need to be computed, while the Bayesian structure is predetermined.

To classify a new instance, Naive Bayes computes the conditional probabilities based on the Bayesian network parameters. It selects the class that maximizes the probability as the predicted class label. Naive Bayes is well-known for its simplicity, yet it has the potential to achieve good predictive performance in various problem domains.

A Multi-Layer Perceptron (MLP) [68] is an artificial neural network commonly used for classification and regression tasks. It enhances the single neuron perceptron by incorporating multiple layers of interconnected artificial neurons. The MLP model consists of an input layer for receiving data features, one or more hidden layers responsible for transforming input signals and adjusting connection weights between neurons, and an output layer that generates the final predictions. To optimize the MLP model, the log-loss function is typically used along with various optimization algorithms such as LBFGS (Limited-memory Broyden- Fletcher-Goldfarb-Shanno) or stochastic Gradient descent. These optimization algorithms aim to search for the optimal set of weights that minimize the loss function, allowing the model to make accurate predictions.

4.2 Evaluation models metrics

There are different metrics to evaluate the accuracy of a model. This is because each metric provides a different perspective on the performance of each algorithm and in some cases avoids biases due to majority classes. The metrics used to evaluate the algorithms used for this study are described below.

AUC is obtained directly from the ROC (Receiver Operating Characteristic) curve, which is a graphical representation of the true positive rate (sensitivity) versus the false positive rate (1 - specificity) for different classification threshold values.

The following metrics are estimated from the confusion matrix, where the true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are placed:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = \frac{2 * TP}{2 * TP + FN + FP} \tag{4}$$

Accuracy is the proportion of modules correctly classified as defective and non- defective. It is calculated by (1).

Precision is the proportion of TP cases. In the SDP, it is the number of defective modules that were found correctly, and it is calculated by (2).

Recall measures the proportion of TP that there should be and is calculated using (3).

Finally, F1-measure is used when there is an unbalanced data set and combines the benefits of precision and recall as shown in (4).

4.3 Classifiers parameters

The implementations of algorithms were carried out in the Python language with the skit- learn 1.4.2 library. Table 5 shows the parameters of the classifiers used for out experiments. In the algorithms that shared some parameter, such as balancing the number of estimators, it was the same for all. The Random state parameter indicates the seed that was the same for all classification algorithms.

5. Results analysis

The average results obtained by the k-fold cross-validation process with k = 10 by each metric were analyzed, both for the data sets without applying any class balancing method and the application of each method described in section 3.

Table 5. Parameters used for each classifier.

Classifier algorithm	**Parameters**
Decision tree	Criterion: Gini Max Depth: None Max leaf nodes: None Min impurity decrease: 0.0 Class weight: None Random state: 42
K-NN	N neighbors: 3 Metric: Euclidean distance
Random Forest	N estimators: 50 Criterion: Gini Max Depth: None Max leaf nodes: None Min impurity decrease: 0.0 Class weight: None Random state: 42
Ada boost	N estimators: 50 Algorithm: SAMME.R Learning rate: 1.0 Random state: 42
Gradient Boosting	N estimators: 50 Loss: log loss Learning rate: 1.0 Criterion: Friedman mse Max depth: 3 Min impurity decrease: 0.0
Naive Bayes	Distribution: Gaussian
Multilayer perceptron	Hidden layer: 5 Random state: 42 Alpha: 0.0005

5.1 Balancing class approaches results

The results show that applying a class balancing method increases the prediction accuracy of different classifiers with different subsets of data through cross-validation, contributes to the increase of reliability in predictions.

With respect to the CM1 data set, performing a random undersampling class balancing procedure worsens the results, while with ALKNN practically the same results are obtained.

For the JM1 data set, all the balancing methods outperform the results without applying any class balancing method except with random undersampling, which obtains very similar results.

Regarding the KC1 data set, most class balancing algorithms outperform the average predictions of using an unbalanced data set. However, using undersampling in the assembled algorithms (AdaBoost and Gradient boosting) results in lower precision values.

All class balancing algorithms outperform the prediction values using the raw KC2 data set.

Finally, the results of the PC1 data set show that the ENN and ALLKNN undersampling algorithms have the worst performance, almost equal to not using a class balancing method.

5.2 Classifiers results

The results show that the classifiers that had the best results in the selected model evaluation metrics are the ensemble classifiers, that is, random forest, AdaBoost and Gradient boosting. The classifier that showed the worst results was the multilayer perceptron. With these results, software engineering practitioners can opt for ensemble classifiers, as it is shown that it is advisable to invest in the processing of several weak classifiers to obtain better prediction accuracy.

5.3 Metrics results

The results of the AUC, accuracy, precision, recall and F1 metrics are similar when the values of the classes of the data sets are balanced. It is quantitatively verified that when the data sets are not balanced, the recall, F1 and precision metrics tend to be lower than accuracy. This is because classifiers tend to predict the majority class, and by performing an evaluation on all the predictions, there is a greater probability that the majority class matches the predicted value. For this reason, it is better to have an overview of different metrics that evaluate accuracy rates differently.

5.4 Statistical Analysis

The metric values are not sufficient to establish a statistically significant improvement in the predictions. Therefore, a suitable statistical test was applied to compare the accuracies among the class balancing models [69].

First, χ^2, Shapiro-Wilk, skewness and kurtosis normality tests for each set of predictions (ten for each metric that were obtained from the cross-validation process) to check if the normality assumption was met were applied.

For each metric and for each data set (set of all classifiers) it was checked if at least some test was not significant at 95%, then a normal distribution was not considered. If all values of each metric had a normal distribution, an ANOVA test was performed, if at least one set did not meet the normality assumption, then Friedman non-parametric test was applied to determine whether at least one algorithm generated significantly different results from the others. The results of the statistical tests can be observed in [70], where it is concluded that in all groups by data set and by class balancing algorithm, there was at least one value statistically different from the rest of them.

Finally, to know if the statistically significant differences were the results related to the data sets without applying any class balancing algorithm, the values of the metrics of each balancing algorithm were compared with the values of not applying a class balancing. To do this, the four normality statistical tests were performed for each set of prediction metrics. If any of their four p-values were lower than 0.05, then data were non-normally distributed at 95% confidence, and a Wilcoxon test was applied (the medians of the models should be compared); otherwise, a t-paired test was performed (the means of models were compared) [71]. The results of these tests can be consulted in [72].

In [72] is shown that the class balancing algorithms for each metric, where the results of the predictions were statistically different from the predictions of the same data set without using any class balancing algorithm at 95% confidence.

These results show that despite different class balancing algorithms and approaches, Naive Bayes and MLP do not generate significantly better predictions. However, using the other classifiers with any metric generally makes a significant difference in the predictions.

6. Validity threats

There are four categories of validity threats in search-based predictive modeling for Software Engineering: conclusion, internal, external and construct [73] of which, we identify the following:

Construct: We remove five records that have null values for the JM1 data set.

Construct: We are not involved in the acquisition of data to avoid bias due to human error.

Internal: Data is based on static code metrics only.

Internal: We did not include any processing to deal with outliers, as we wanted to evaluate them with real data from the PROMISE repository systems.

7. Conclusions and future work

The present study experiments with different approaches to mitigate the class balancing problem in SDP. These approaches are obtained from a literature review, of which three oversampling algorithms, four undersampling algorithms and two from the combination of the mentioned approaches were selected. The data sets used were those from the well-known public repository PROMISE, to obtain predictions with the data used in software industry.

To not bias the results with a particular classifier, seven classification algorithms (based on different approaches) were tested. Also, different metrics such as AUC, precision, accuracy, recall and F1 were evaluated to measure the performance of the algorithms with and without class balancing through a cross- validation process. Finally, the results are statistically evaluated to conclude if there really is a significant improvement when balancing the data sets.

With respect to the class balancing methods, we conclude that:

- Balancing the classes significantly improves prediction performance since learning is not biased by the majority class.
- The oversampling and combination methods have better performance than the algorithms based on undersampling.

With respect to the classification algorithms, we conclude that:

- The algorithms that obtained the best results were the assembled classifiers (both boosting and bagging), which base the predicted class value on the results of several weak classifiers. Therefore, if software engineering practitioners require greater precision, it is important to invest resources in these types of methods.
- The multilayer perceptron and Naïve Bayes algorithms are the classifier that yielded the worst results.

In terms of evaluation metrics, we conclude that:

- When the data sets are unbalanced, the metrics show very variable results.
- When the data sets are balanced, the results of the metrics are more homogeneous.

Finally, it is concluded that to improve the reliability of a software product, it is necessary to have good precision in predictions. Therefore, when the historical data to make predictions of software defects is unbalanced, it is suggested to first apply an oversampling mechanism and then use an ensemble classifier that strengthens the predictions, regardless of the set used for training and testing.

As future work we propose the following:

Repeat the experimentation with other classifiers including deep learning, since we only chose some representative ones of each approach.

As could be seen in Table 5, the classification algorithms need to adjust parameters, so it is proposed to implement some optimization algorithm that allows knowing the best parameters for each classification algorithm.

It is also proposed to carry out this analysis by programming language [74], since the development language could influence the introduction of defects and the quality of the software [75].

Finally, continue with experimentation on other data sets, which allows exploring the performance of class balancing algorithms in more dept.

References

[1]. D. J. Olvera-Villeda, A. J. Sánchez-García, X. Limón, and S. Domínguez Isidro, "Class balancing approaches in dataset for software defect prediction: A systematic literature review" in 2023 11th International Conference in Software Engineering Research and Innovation (CONISOFT). IEEE, 2023, pp. 1–6.

[2]. M. Glinz, "A glossary of requirements engineering terminology", Standard Glossary of the Certified Professionalfor Requirements Engineering (CPRE) Studies and Exam, Version, vol. 1, p. 56, 2011.

[3]. J. D. Musa, "Software reliability measurement", Journal of Systems and Software, vol. 1, pp. 223–241, 1979.

[4]. I. Iso and N. IEC, "Iso/iec", IEEE International Standard-Systems and software engineering- Vocabulary, pp. 1–541,2017.

[5]. P. D. Singh and A. Chug, "Software defect prediction analysis using machine learning algorithms" in 2017 7th international conference on cloud computing, data science & engineering confluence. IEEE, 2017, pp. 775–781.

[6]. J. Sayyad Shirabad and T. Menzies, "The PROMISE Repository of Software Engineering Databases", School of Information Technology and Engineering, University of Ottawa, Canada, 2005. [Online]. Available: http://PROMISEsite.uottawa.ca/SERepository

[7]. T. McCabe, "A complexity measure", IEEE Transactions on Software Engineering, vol. 2, no. 4, pp. 308–320, December 1976.

[8]. M. Halstead, Elements of Software Science. Elsevier, 1977.

[9]. D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization", IEEE transactions on evolutionary computation, vol. 1, no. 1, pp. 67–82, 1997.

[10]. Y. Zhang, X. Yan, and A. A. Khan, "A kernel density estimation-based variation sampling for class imbalance in defect prediction" in 2020 IEEE Intl Conf. on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom). IEEE, 2020, pp. 1058–1065.

[11]. E. Elahi, S. Kanwal, and A. N. Asif, "A new ensemble approach for software fault prediction" in 2020 17th international Bhurban conference on applied sciences and technology (IBCAST). IEEE, 2020, pp. 407–412.

[12]. J. Zheng, X. Wang, D. Wei, B. Chen, and Y. Shao, "A novel imbalanced ensemble learning in software defect predication", IEEE Access, vol. 9, pp. 86 855–86 868, 2021.

[13]. Q. Zha, X. Yan, and Y. Zhou, "Adaptive centre-weighted oversampling for class imbalance in software defect prediction", in 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom). IEEE, 2018, pp. 223–230.

[14]. S. Huda, K. Liu, M. Abdelrazek, A. Ibrahim, S. Alyahya, H. Al-Dossari, and S. Ahmad, "An ensemble oversampling model for class imbalance problem in software defect prediction", IEEE access, vol. 6, pp. 24 184–24 195, 2018.

[15]. R. Malhotra, N. Nishant, S. Gurha, and V. Rathi, "Application of particle swarm optimization for software defect prediction using object oriented metrics" in 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2021, pp. 88–93.

[16]. Z. Li, X. Zhang, J. Guo, and Y. Shang, "Class imbalance data generation for software defect prediction", in 2019 26th Asia-Pacific Software Engineering Conference (APSEC). IEEE, 2019, pp. 276–283.

[17]. S. Ghosh, A. Rana, and V. Kansal, "Combining integrated sampling with nonlinear manifold detection techniques for software defect prediction" in 2018 3rd International Conference on Contemporary Computing and Informatics (IC3I). IEEE, 2018, pp. 147–154.

[18]. S. A. Putri et al., “Combining integreted sampling technique with feature selection for software defect prediction” in 2017 5th International Conference on Cyber and IT Service Management (CITSM). IEEE, 2017, pp. 1–6.

[19]. T. Thaher and N. Arman, “Efficient Multi-Swarm Binary Harris Hawks Optimization as a Feature Selection Approach for Software Fault Prediction” in 2020 11th International conference on information and communication systems (ICICS). IEEE, 2020, pp. 249–254.

[20]. K. Bashir, T. Li, C. W. Yohannese, and Y. Mahama, “Enhancing software defect prediction using supervised-learning based framework” in 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). IEEE, 2017, pp. 1–6.

[21]. S. S. Rathore, S. S. Chouhan, D. K. Jain, and A. G. Vachhani, “Generative oversampling methods for handling imbalanced data in software fault prediction”, IEEE Transactions on Reliability, vol. 71, no. 2, pp. 747–762, 2022.

[22]. Z. Eivazpour and M. R. Keyvanpour, “Improving performance in software defect prediction using variational autoencoder” in 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI). IEEE, 2019, pp. 644–649.

[23]. A. Bispo, R. Prudˆencio, and D. VÅLeras, “Instance selection and class balancing techniques for cross project defect prediction” in 2018 7th Brazilian Conference on Intelligent Systems (BRACIS). IEEE, 2018, pp. 552–557.

[24]. K. E. Bennin, J. Keung, P. Phannachitta, A. Monden, and S. Mensah, “Mahakil: Diversity based oversampling approach to alleviate the class imbalance issue in software defect prediction” IEEE Transactions on Software Engineering, vol. 44, no. 6, pp. 534–550, 2017.

[25]. R. Malhotra, R. Kapoor, P. Saxena, and P. Sharma, “Saga: A hybrid technique to handle imbalance data in software defect prediction” in 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). IEEE, 2021, pp. 331–336.

[26]. D. Wang and X. Xiong, “Software defect prediction based on combined sampling and feature selection” in ICMLCA 2021; 2nd International Conference on Machine Learning and Computer Application. VDE, 2021, pp. 1–5.

[27]. Y. Liu, F. Sun, J. Yang, and D. Zhou, “Software defect prediction model based on improved bp neural network” in 2019 6th International Conference on Dependable Systems and Their Applications (DSA). IEEE, 2020, pp. 521–522.

[28]. R. B. Bahaweres, F. Agustian, I. Hermadi, A. I. Suroso, and Y. Arkeman, “Software defect prediction using neural network based smote” in 2020 7th International Conference on Electrical Engineering, Computer Sciences and Informatics (EECSI). IEEE, 2020, pp. 71–76.

[29]. S. Choirunnisa, B. Meidyani, and S. Rochimah, “Software defect prediction using oversampling algorithm: A-suwo” in 2018 Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS). IEEE, 2018, pp. 337– 341.

[30]. W. A. Dipa and W. D. Sunindyo, “Software defect prediction using smote and artificial neural network” in 2021 International Conference on Data and Software Engineering (ICoDSE). IEEE, 2021, pp. 1– 4.

[31]. R. Malhotra, V. Agrawal, V. Pal, and T. Agarwal, “Support vector based oversampling technique for handling class imbalance in software defect prediction” in 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2021, pp. 1078–1083.

[32]. L. Gong, S. Jiang, and L. Jiang, “Tackling class imbalance problem in software defect prediction through clusterbased over-sampling with filtering” IEEE Access, vol. 7, pp. 145 725–145 737, 2019.

[33]. R. Malhotra and S. Kamal, “Tool to handle imbalancing problem in software defect prediction using oversampling methods” in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2017, pp. 906–912.

[34]. S. K. Pandey and A. K. Tripathi, “Class imbalance issue in software defect prediction models by various machine learning techniques: an empirical study” in 2021 8th International Conference on Smart Computing and Communications (ICSCC). IEEE, 2021, pp. 58–63.

[35]. W. Zhang, Y. Li, M. Wen, and R. He, “Comparative study of ensemble learning methods in just-in-time software defect prediction” in 2023 IEEE 23rd International Conference on Software Quality, Reliability, and Security Companion (QRSC), 2023, pp. 83–92.

[36]. X. Yang, S. Wang, Y. Li, and S. Wang, “Does data sampling improve deep learning-based vulnerability detection? yeas! and nays!” in 2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE), 2023, pp. 2287–2298.

[37]. R. Kumar and A. Chaturvedi, “Software bug prediction using reward-based weighted majority voting ensemble technique”, IEEE Transactions on Reliability, vol. 73, no. 1, pp. 726–740, 2024.

[38]. M. Devi, T. Rajkumar, and D. Balakrishnan, "Prediction of software defects by employing optimized deep learning and oversampling approaches" in 2024 2nd International Conference on Computer, Communication and Control (IC4), 2024, pp. 1–5.

[39]. W. Wei, F. Jiang, X. Yu, and J. Du, "An under- sampling algorithm based on weighted complexity and its application in software defect prediction" in Proceedings of the 2022 5th International Conference on Software Engineering and Information Management, 2022, pp. 38–44.

[40]. G. Abaei, W. Z. Tah, J. Z. W. Toh, and E. S. J. Hor, "Improving software fault prediction in imbalanced datasets using the under-sampling approach" in Proceedings of the 2022 11th International Conference on Software and Computer Applications, 2022, pp. 41–47.

[41]. Z.-W. Zhang, X.-Y. Jing, and T.-J. Wang, "Label propagation based semi-supervised learning for software defect prediction", Automated Software Engineering, vol. 24, pp. 47–69, 2017.

[42]. X. Du, H. Yue, and H. Dong, "Software defect prediction method based on hybrid sampling" in International Conference on Frontiers of Electronics, Information and Computation Technologies, ser. ICFEICT 2021. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3474198.3478215.

[43]. D. Ryu, J.-I. Jang, and J. Baik, "A transfer cost- sensitive boosting approach for cross-project defect prediction", Software Quality Journal, vol. 25, pp. 235–272, 2017.

[44]. L. Zhou, R. Li, S. Zhang, and H. Wang, "Imbalanced data processing model for software defect prediction", Wireless Personal Communications, vol. 102, pp. 937–950, 2018.

[45]. H. He, X. Zhang, Q. Wang, J. Ren, J. Liu, X. Zhao, and Y. Cheng, "Ensemble multiboost based on ripper classifier for prediction of imbalanced software defect data", IEEE Access, vol. 7, pp. 110 333–110 343, 2019.

[46]. C. Zeng, C. Y. Zhou, S. K. Lv, P. He, and J. Huang, "Gcn2defect: Graph convolutional networks for smotetomek based software defect prediction" in 2021 IEEE 32nd International Symposium on Software Reliability Engineering (ISSRE). IEEE, 2021, pp. 69–79.

[47]. A. Joon, R. K. Tyagi, and K. Kumar, "Noise filtering and imbalance class distribution removal for optimizing software fault prediction using best software metrics suite" in 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2020, pp. 1381–1389.

[48]. L. Chen, B. Fang, Z. Shang, and Y. Tang, "Tackling class overlap and imbalance problems in software defect prediction", Software Quality Journal, vol. 26, pp. 97–125, 2018.

[49]. S. Riaz, A. Arshad, and L. Jiao, "Rough noise-filtered easy ensemble for software fault prediction", IEEE Access, vol. 6, pp. 46 886–46 899, 2018.

[50]. X. Wan, Z. Zheng, and Y. Liu, "Spe2: Self-paced ensemble of ensembles for software defect prediction", IEEE Transactions on Reliability, vol. 71, no. 2, pp. 865–879, 2022.

[51]. G. Menardi and N. Torelli, "Training and assessing classification rules with imbalanced data", Data Mining and Knowledge Discovery, vol. 28, no. 1, pp. 92–122, 2012. [Online]. Available: http://dx.doi.org/10.1007/s10618-012-0295-5.

[52]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-samplingtechnique", Journal of Artificial Intelligence Research, vol. 16, no. nil, pp. 321–357, 2002. [Online]. Available: http://dx.doi.org/10.1613/jair.953.

[53]. H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning" in 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 6 2008, p. nil. [Online]. Available: http://dx.doi.org/10.1109/IJCNN.2008.4633969.

[54]. G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data", ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 20–29, 2004. [Online]. Available: http://dx.doi.org/10.1145/1007730.1007735.

[55]. I. Mani and I. Zhang, "knn approach to unbalanced data distributions: a case study involving information extraction" in Proceedings of workshop on learning from imbalanced datasets, vol. 126, no. 1. ICML, 2003, pp. 1–7.

[56]. D. L. Wilson, "Asymptotic properties of nearest neighbor rules using edited data", IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-2, no. 3, pp. 408–421, 1972.

[57]. I. Tomek, "An experiment with the edited nearest- neighbor rule", IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-6, no. 6, pp. 448–452, 1976. [Online]. Available: http://dx.doi.org/10.1109/TSMC.1976.4309523.

[58]. B. R. Manju and A. R. Nair, “Classification of Cardiac Arrhythmia of 12 Lead ECG Using Combination of SMOTEENN, XGBoost and Machine Learning Algorithms” in 2019 9th International Symposium on Embedded Computing and System Design (ISED), 2019, pp. 1–7.
[59]. G. E. A. P. A. Batista, A. L. C. Bazzan, and M. C. Monard, “Balancing training data for automated annotation of keywords: a case study” in WOB, 2003. [Online]. Available: https://api.semanticscholar.org/CorpusID:1579194
[60]. I. Tomek, “Two modifications of cnn”, IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-6, no. 11, pp. 769–772, 1976.
[61]. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees, ser. Routledge, 2017. [Online]. Available: http://dx.doi.org/10.1201/9781315139470.
[62]. D. A. Cieslak and N. V. Chawla, “Learning decision trees for unbalanced data” in Machine Learning and Knowledge Discovery in Databases, W. Daelemans, B. Goethals, and K. Morik, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 241–256.
[63]. E. Fix and J. L. Hodges, “Discriminatory analysis. nonparametric discrimination: Consistency properties”, International Statistical Review / Revue Internationale de Statistique, vol. 57, no. 3, p. 238, 1989. [Online]. Available: http://dx.doi.org/10.2307/1403797.
[64]. L. Breiman, “Random forests”, Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: http://dx.doi.org/ 10.1023/A:1010933404324
[65]. Y. Freund and R. E. Schapire, “Experiments with a new boosting algorithm” in Proceedings of the Thirteenth International Conference on International Conference on Machine Learning, ser. ICML’96. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1996, p. 148–156.
[66]. J. H. Friedman, “Stochastic Gradient boosting”, Computational Statistics Data Analysis, vol. 38, no. 4, pp. 367–378, 2002, nonlinear Methods and Data Mining. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S01 67947301000652.
[67]. D. J. Hand and K. Yu, “Idiot’s bayes-not so stupid after all?” International Statistical Review, vol. 69, no. 3, pp. 385–398, 2001. [Online]. Available: http://dx.doi.org/10.1111/j.1751-823.2001.tb00465.x.
[68]. G. E. Hinton, Connectionist learning procedures, ser. Machine Learning. Elsevier, 1990, pp. 555–610. [Online]. Available: http://dx.doi.org/10.1016/B978- 0-08-051055-2.50029-8.
[69]. T. Dyba, V. B. Kampenes, and D. I. Sj.berg, “A systematic review of statistical power in software engineering experiments”, Information and Software Technology, vol. 48, no. 8, pp. 745–755, 2006.
[70]. J. Sánchez-García, “Statistical tests among groups”, [Data set], 2024. [Online]. Available: https://doi.org/10.5281/zenodo.13239734.
[71]. D. S. Moore and G. P. McCabe, Introduction to the practice of statistics. WH Freeman/Times Books/Henry Holt & Co, 1989.
[72]. J. Sánchez-García, “Statistical tests results”, [Data set], 2024. [Online]. Available: https://doi.org/10.5281/zenodo.13240040.
[73]. R. Malhotra and M. Khanna, “Threats to validity in search based predictive modelling for software engineering”, IET Software, vol. 12, no. 4, pp. 293-305, 2018.
[74]. I. Bronshteyn, “Study of defects in a program code in python”, Programming and Computer Software, vol. 39, pp. 279–284, 2013.
[75]. A. Belevantsev, “Multilevel static analysis for improving program quality”, Programming and Computer Software, vol. 43, pp. 321–336, 2017.

Информация об авторах / Information about authors

Ангел Хуан САНЧЕС-ГАРСИЯ имеет степень PhD по искусственному интеллекту, доцент факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных интересов: программометрия, машинное обучение, прогнозирование затрат, эволюционные вычисления.

Ángel Juan SÁNCHEZ-GARCÍA – doctor in artificial intelligence, associate professor at Facultad de Estadística e Informática, Universidad Veracruzana, Mexico (School of Statistics and Informatics, University of Veracruz). Research interests: software measurement, machine learning, effort prediction, evolutionary computation.

Рианьо Гектор Ксавьер ЛИМОН имеет степень PhD по искусственному интеллекту, доцент факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных

интересов: интеллектуальный анализ данных, мультиагентные и распределенные системы, архитектура программного обеспечения.

Riaño Hector Xavier LIMÓN – doctor in artificial intelligence, associate professor at Facultad de Estadística e Informática, Universidad Veracruzana, Mexico (School of Statistics and Informatics, University of Veracruz). Research interests: data mining, multi-agent systems, distributed systems, and software architecture.

Саул ДОМИНГЕС-ИСИДРО имеет степень PhD по искусственному интеллекту, доцент факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных интересов: распределенные системы, разработка программного обеспечения, вычислительный интеллект, машинное обучение.

Saúl DOMÍNGUEZ-ISIDRO – PhD in artificial intelligence, associate professor at Facultad de Estadística e Informática, Universidad Veracruzana, Mexico (School of Statistics and Informatics, University of Veracruz). Research interests: distributed systems, software development, computational intelligence, and machine learning.

Дан Хавьер ОЛВЕРА-ВИЙЕДА имеет степень бакалавра в области программной инженерии, сотрудник факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных интересов: разработка программного обеспечения, вычислительный интеллект, машинное обучение.

Dan Javier OLVERA-VILLEDA – bachelor in software engineering, Facultad de Estadística e Informática, Universidad Veracruzana, Mexico (School of Statistics and Informatics, University of Veracruz). Research interests: software development, machine learning, and computational intelligence.

Хуан Карлос ПЕРЕС-АРРИАГА имеет степень магистра по программированию, доцент факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных интересов: доступность программного обеспечения, безопасность программного обеспечения, конструирование программ.

Juan Carlos PÉREZ-ARRIAGA has a master’s degree in computer science, an associate professor at Facultad de Estadística e Informática, Universidad Veracruzana, Mexico (School of Statistics and Informatics, University of Veracruz). Research interests: software accessibility, software security, and software construction.

DOI: 10.15514/ISPRAS-2024-36(6)-3

Domain-Driven Design in Microservices Architecture

J. Sangabriel-Alarcón, ORCID: 0009-0002-2682-502X <josusangabriel@uv.mx>
J. O. Ocharán-Hernández, ORCID: 0000-0002-2598-1445 <jocharan@uv.mx>
X. Limón, ORCID: 0000-0003-4654-636X <hlimon@uv.mx>
M. K. Cortés-Verdín, ORCID: 0000-0002-6453-180X <kcortes@uv.mx>
School of Statistics and Informatics, Universidad Veracruzana, Xalapa, Veracruz, México.

Abstract. With the increment in software development complexity, approaches such as Domain-Driven Design (DDD) are needed to tackle contemporary business domains. DDD is already being used in various software projects with different architectural styles. Although some studies have explored the decomposition of business domains or legacy monolithic systems into microservices, there is a lack of concrete information regarding the practical implementation of DDD in this architectural style. The paper systematizes findings on the purpose of using DDD, its patterns, associated technologies, and techniques to increase the clarity about the use of DDD in microservices-based systems development. A systematic literature review of 35 articles was conducted. Thematic analysis was employed to identify five high-order themes and 11 themes. Based on our analysis, we have concluded that microservice identification emerges as the primary motivation behind developers' adoption of DDD, but not the only usage of DDD reported in the literature. Finally, our analysis found benefits and challenges in the use of DDD in Microservices Architecture which are translated to opportunity areas for future works.

Keywords: domain-driven design; microservices architecture; systematic literature review; thematic analysis.

For citation: Sangabriel-Alarcón J., Ocharán-Hernández J. O., Limón X., Cortés-Verdín M. K. Domain-Driven Design in Microservices Architecture. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 39-58. DOI: 10.15514/ISPRAS-2024-36(6)-3.

Acknowledgements. This work was supported by ongoing institutional funding. No additional grants to carry out or direct this particular research were obtained.

Предметно-ориентированное проектирование в микросервисной архитектуре

Х. Сангабриэль-Аларкон, ORCID: 0009-0002-2682-502X <josusangabriel@uv.mx>
Х. О. Очаран-Эрнандес, ORCID: 0000-0002-2598-1445 <jocharan@uv.mx>
К. Лимон, ORCID: 0000-0003-4654-636X <hlimon@uv.mx>
М. К. Кортес-Вердин, ORCID: 0000-0002-6453-180X <kcortes@uv.mx>
Школа статистики и информатики, Университет Веракруса, Халапа, Веракрус, Мексика.

Аннотация. С увеличением сложности разработки программного обеспечения для решения современных бизнес-задач необходимы такие подходы, как предметно-ориентированное проектирование (Domain-Driven Design, DDD). DDD уже используется в различных программных проектах с разными архитектурными стилями. Хотя в некоторых исследованиях изучалось разложение бизнес-доменов или унаследованных монолитных систем на микросервисы, пока отсутствует конкретная информация относительно практической реализации DDD в этом архитектурном стиле. Для повышения ясности в отношении использования DDD в разработке систем на основе микросервисов в нашей статье систематизированы выводы о целях использования DDD, его моделях, связанных технологиях и методах. Нами был проведен систематический обзор литературы из 35 статей. Тематический анализ помог выявить 11 тем и пять тем более высокого порядка. Основываясь на проделанном анализе, мы пришли к выводу, что идентификация микросервисов становится основной мотивацией принятия разработчиками DDD, но при этом вовсе не является единственной причиной использования DDD, о которой сообщается в литературе. Наконец, наш анализ выявил преимущества и проблемы в использовании DDD в архитектуре микросервисов, которые будут учитываться при проведении работ в будущем.

Ключевые слова: предметно-ориентированное проектирование; микросервисная архитектура; систематический литературный обзор; тематический анализ.

Для цитирования: Сангабриэль-Аларкон Х., Очаран-Эрнандес Х. О., Лимон К., Кортес-Вердин М. К. Предметно-ориентированное проектирование в микросервисной архитектуре. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 39–58 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–3.

Благодарности. Работа была поддержана текущим институциональным финансированием. Никаких дополнительных грантов для проведения или руководства этим конкретным исследованием получено не было.

1. Introduction

This paper is an extension of work initially presented at the 11th International Conference in Software Engineering Research and Innovation (CONISOFT 2023) [1]. The original study is a systematic mapping study on Domain-Driven Design (DDDS) for Microservices Architecture Systems Development. In this paper, we conducted a comprehensive systematic literature review and employed thematic synthesis to identify and analyze patterns in the uses of DDD in this context. This study synthesizes the findings from a broader range of primary studies and search strategies.

Since the release of Eric Evans' book "The Blue Book" in 2004 [2], a community of practitioners has emerged who explore the use of DDD and patterns in different software development projects. DDD can be understood as an approach that addresses the complexities of a business by emphasizing the team's focus on domain knowledge [2]. Some authors [3-5] have proposed patterns and techniques to analyze business domains and incorporate that knowledge into software projects.

DDD patterns can be classified as strategic and tactical designs, which are the key elements of this approach. Strategic design involves domain analysis and decomposition. On the other hand, tactical design translates the knowledge acquired from strategic design into actual lines of code [4].

When applying MSA in practice, developers have encountered a range of challenges in achieving the desired properties of this architectural style [6-9]. Based on the challenges highlighted by various

authors concerning MSA, several studies aim to reduce the complexity in the development of microservices-based systems. While some proposals have been put forth to address these challenges, no definitive solution has emerged. However, the use of DDD has been featured prominently in these proposals. This connection between DDD and MSA can be traced back to the 2014 definition of MSA, and the principles of DDD have frequently been referenced within microservices.

Despite the theoretical discussion [10-11] of the relationship between these approaches, uncertainties have surrounded their practical application. Additionally, the challenges associated with MSA [7, 9], combined with those of DDD [11], have given rise to new challenges in the practical implementation of microservices development with DDD. This lack of clarity regarding DDD for microservices development has created a gap between theory and practice. The issue concerning the knowledge gap between the theory and practice of DDD is directly tied to the practical advantages it offers in designing microservices-based systems. In some studies, [12-13], a set of strategies and techniques to design APIs have been reported to deal important decisions about microservices, where the granularity definition of microservices has been one of the most important themes around mentioned architecture. However, various ideas, patterns, and techniques within DDD have often been cited concerning achieving mentioned benefits related to well-defined granularity for microservices. Moreover, incorporating DDD principles suggests enhanced efficiency in stakeholder collaborative work [13]. The mentioned benefits of DDD are examples of potential solutions for microservices challenges, which are hindered by the lack of knowledge about the use of DDD principles and patterns in a practical context.

This lack of clarity can be mitigated by analyzing practical cases documented in the literature regarding the utilization of DDD in the development of microservices-based systems. While other studies have addressed this issue, none provide specific reasons for authors' decisions to employ DDD in microservices development and the corresponding outcomes. Among these unexplored aspects, it is crucial to determine the DDD techniques employed, identify the patterns utilized in microservices design, and explore other pertinent details that shed light on the limitations and areas of opportunity in using DDD for microservices-based development.

To better understand the practical use of DDD in developing microservices-based systems, this study extends our previously systematic mapping study [1]. We conducted a systematic literature review complemented with a thematic analysis. The objective was to investigate the current state of the art in microservices development with DDD in practical scenarios. We compiled and analyzed a collection of diverse studies that reported the use of DDD in their microservices projects from 2014 to 2023. The findings obtained from this evidence can assist developers in identifying the practical applications and adaptations made by their peers when utilizing DDD. Furthermore, this research can also help uncover DDD's limitations and identify areas of opportunity where this approach can effectively address the main challenges associated with microservices development.

This study is organized as follows: Section 2 provides an overview of the related work in this field. Section 3 outlines the research method chosen for conducting the systematic literature review. The execution of this research method is described in Section 4. The results obtained from the research method are presented in Section 5, followed by a discussion in Section 6. Section 7 addresses the potential threats to the validity of this study. Finally, Section 8 presents the conclusions drawn from the evidence collected in this research.

2. Related work

This section provides an overview of the research work associated with the objective of this study. In a study by Singjai et al. 2021 [14], the authors investigated Architectural Design Decisions (ADD) associated with API design and DDD patterns using a grounded theory research method. The APIs developed using DDD served as the foundation for modeling microservices. Specifically, Singjai et al. identified six ADDs and 27 decision options about utilizing the DDD domain model and strategic patterns for delineating API specifics. It is important to note that this study was limited to gray

literature and did not analyze white literature sources. Singjai et al. acknowledged the risk of generalizing their findings, underscoring the potential for additional resources in different data sources relevant to the research topic.

Schmidt et al. also conducted a relevant study in 2020 [15], which entailed a systematic literature review focused on microservices identification proposals. This study examined two distinct development approaches: Model-Driven Development (MDD) and DDD. The authors collected a set of primary studies from 2013 to 2019, of which 27 were considered for review. Among these 27 primary studies, only four included DDD patterns specifically for microservices identification. While the study primarily focused on white literature, it did not specifically address the examination of DDD and MSA practices as its main objective. Given the time frame covered by Schmidt et al. and the recent surge in research highlighting the relationship between DDD and MSA, as mentioned in various studies [16-18], there arises a clear need for a dedicated study that concentrates on the utilization of DDD within the context of developing microservices-based systems.

This study will solely encompass white literature to narrow the research scope and delve into previously unexplored evidence of the integration of DDD and MSA. By concentrating solely on white literature, we aim to complement existing related work and provide an analysis of DDD's application in developing microservices-based systems.

3. Research method and conduction

This section describes the method followed for our systematic literature review. Firstly, we followed the Kitchenham proposal [19] for evidence-based research on software engineering. In addition, other methods were selected to complement some phases and activities of the research. The methods used to complement the systematic literature review were: (I) Automatic search with Zhang et al. proposal [20], (II) Snowballing process proposed by Wohlin [21], (III) Narrative synthesis from Popay et al. proposal [22], and (IV) Thematic synthesis from the proposal of Cruzes & Dyba [23]. To complement the analysis with thematic synthesis, some guidelines of the Thematic analysis method proposed by Clark & Braun [24] were performed in this study.

3.1 Search process

We started following Phase 1 of the systematic literature review method proposed by Kitchenham. Following the mentioned method, we defined and refined a set of research questions (RQ) during the research process. These RQs were documented in a systematic literature review protocol [25] and uploaded in Zenodo [25].

These RQs were guidelines for the research process and the key criteria for discarding or selecting papers during the search process. Through a manual search, relevant studies were identified, and they were the basis for performing an automatic search. We chose the proposal from Zhang et al. [20] to create a search string that facilitates the identification of primary studies on different engines. The automatic search method of Zhang et al. is closely related to systematic literature review studies [19], and the proposal of strict metrics to evaluate the quality of a search string (Recall and Precision) reduces the likelihood of missing relevant studies.

Following the automatic search proposal by Zhang et al., the relevant studies found from manual search formed the Quasi-Gold Standard (QGS) [20]. The relevant studies identified after manual search were published in the following databases: IEEE Xplore, ACM Digital Library, ScienceDirect, and SpringerLink. With 18 studies that conformed to the QGS and IEEE Xplore selected as the evaluation engine, we evaluated different versions of the search string with the Recall and Precision metrics following the recommendations of Zhang et al. [20], where the search string was only approved when its recall was at least 80%. Several iterations of search string evaluation were performed.

3.2 Selection process

Once the search string was built, we established inclusion and exclusion criteria based on the characteristics of primary studies that formed the QGS. For the filter process, the selection criteria were grouped into four stages. Stage 1 was performed through the year and type filters of engines selected. Stage 2 grouped exclusion criteria related to the access of papers and the duplicated studies between engines. This duplication was identified mainly between ACM Digital Library and IEEE Xplore, where four duplicated studies were found during the manual search. Stage 3 involved the reading of the title and abstract of each paper. Lastly, in Stage 4, the papers were downloaded and read to confirm that the content answered at least one RQ.

As a result of the selection process, some papers were included and excluded through different stages. A sum of 624 studies was collected from the execution of search strings in all engines. After Stage 1, 357 studies were filtered. In Stage 2, 155 studies were discarded. After Stage 3, 79 studies were discarded, obtaining 123 relevant studies. As a result of Stage 4, 31 primary studies were identified.

3.3 Snowballing process

After the selection process, 31 primary studies were identified. However, some primary studies could have been omitted during the selection process, so we decided to perform a snowballing process. We chose the process proposed by Wohlim [21]. This method proposes a systematic selection based on the relationship between studies through their references, which allows the division of the entire process into backward and forward. Firstly, we performed a backward snowballing, followed by a foreward snowballing. At the end of the snowballing process, 35 primary studies were identified as sum of the primary studies of automatic search and the four primary studies found in snowballing [26-60].

3.4 Data extraction process

Following the recommendations of Kitchenham for a systematic literature review, we performed a preliminary synthesis based on the proposal of Popay et al. [22] to identify the answers to the RQs. We performed only some steps of narrative synthesis to confirm that each primary study answered at least one RQ. This preliminary synthesis also allowed us to familiarize ourselves with the content of primary studies. This familiarization phase is one of the first steps of thematic synthesis [23]. We also performed a thematic synthesis, where the data was combined, and grouped into Themes to express higher-order ideas such as Cruzes and Dyba expressed in their proposal [23].

3.5 Data synthesis

As part of Phase 2 of Kitchenham's method is the Synthesis of research. This process is a crucial part of the analysis of evidence. Through an interpretative and systematic process, new knowledge is generated based on a set of data. Thematic synthesis allows us to combine, compare, and explore the patterns in the data. These meaning patterns are helpful in generating new conclusions (generalizations) to achieve the aim of this study and complete Kitchenham's method. The thematic synthesis method was based on the thematic analysis proposed by Braun and Clark [24], providing guidelines to explore the evidence and cover the step "Data synthesis" of the Kitchenham method. The first step of thematic synthesis corresponds to familiarization with primary studies. The mentioned preliminary synthesis was also used to cover the first step. The second is identifying text segments from primary studies that answer the RQs. This second step inspired the second data extraction for thematic synthesis mentioned in Section 3.4. The third step was the label of text segments. This step was performed through coding in MaxQDA[1] 2020, where the codes were filled

[1] https://www.maxqda.com/

out in the tabular formats described in Section 3.4. The code labels were refined through consensus among the authors of this study.

The fourth step was the identification of themes as a set of closely related codes. Each code represents a relevant and single-faceted concept, while each theme represents a multi-faceted idea [24]. Therefore, we grouped codes that explain the meaning of the same background idea. After identifying the themes, they were grouped into higher-level taxonomies (Higher-order themes). These higher-order themes related to a set of themes show the high-level overview of the data from the evidence collected.

4. Results

The products obtained from the method conduction are shown in this section. These results encompass answers for RQs and a thematic map that synthesizes all data collected by this study. The use of DDD in microservices-based systems development has increased in recent years, where 2023 represents the year with the most significant number of primary studies published. The distribution graphic with the publication years of primary studies is shown in Fig. 1.

Regarding the engines where primary studies were published, the IEEE Xplore was the engine where the major number of primary studies were found, with 25 primary studies published. ACM Digital Library was the second, with five primary studies published, ScienceDirect with three studies, and SpringerLink with two studies. The graphic with the number of primary studies found per engine is shown in Fig. 2.

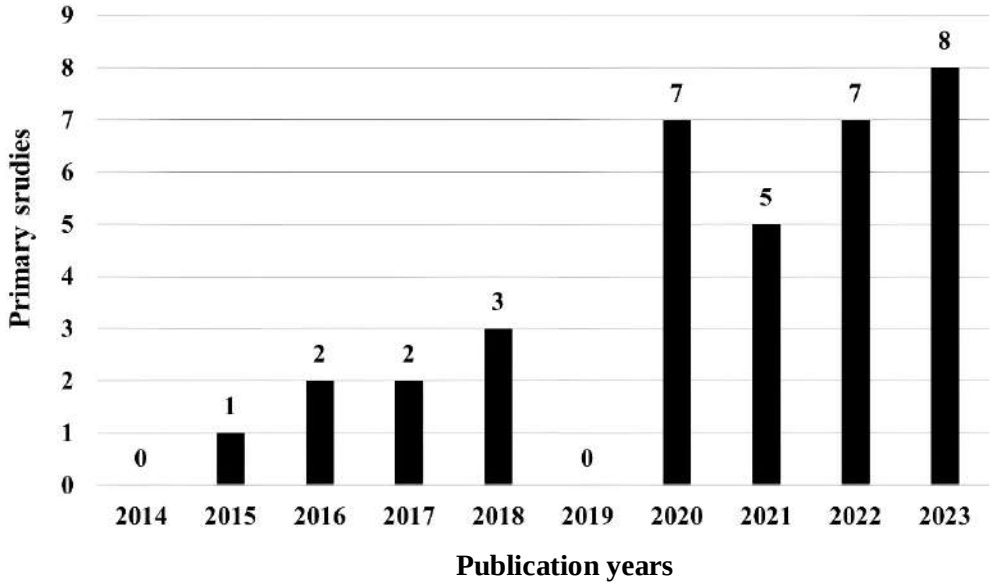


Fig. 1. Primary studies by publication years.

4.1 Answers to research questions

The RQs were the guidelines of this study, and the findings obtained for each RQ enabled us to understand the state of research on the use of DDD for microservices-based systems. This section answers the research questions mainly with quantitative data and some qualitative details. However, the product of qualitative analysis is shown in Section 5.2.

(1) RQ-1: What are the purposes of using DDD for microservices-based systems development? In the use of DDD reported by authors, four motivations were identified in microservices-based systems development. These motivations are shown in Fig. 3.

As shown in Fig. 3, almost all authors of the primary studies mentioned having used DDD for microservices identification [26-52, 54-60], which consists of decomposing a business domain or legacy system into partitions corresponding to microservice candidates.

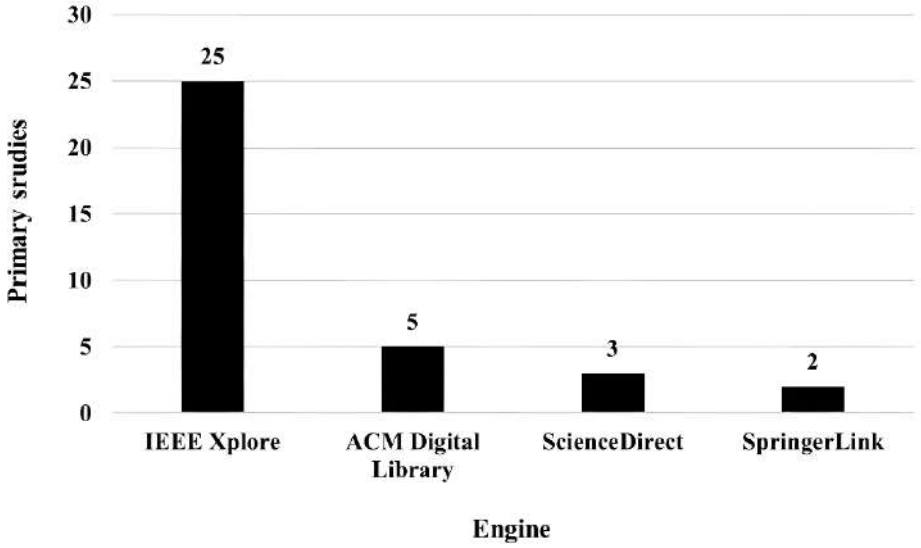


Fig. 2. Primary studies by engine.

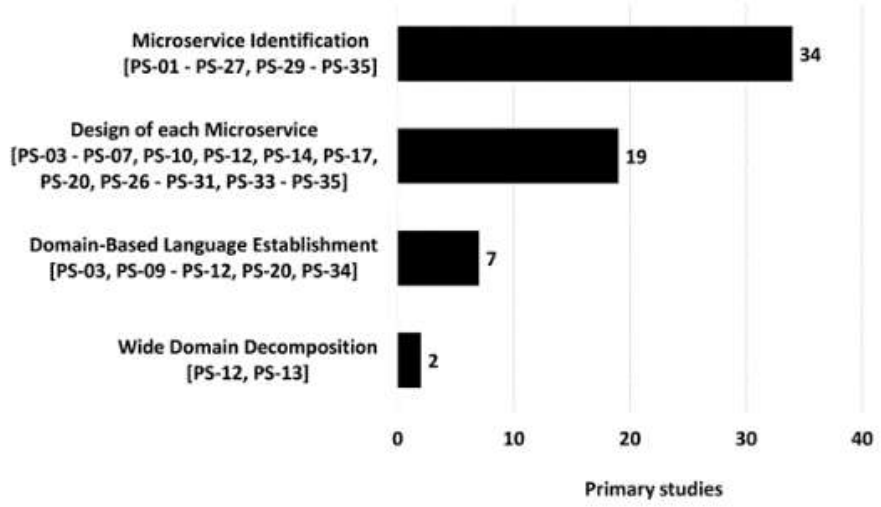


Fig. 3. Purposes of DDD Usage in Microservices-Based Systems Development.

For this purpose, the strategic patterns BC and Subdomain were commonly used. The second most frequent purpose by which authors chose DDD was the design of each microservice [28-32, 35, 37, 39, 42, 45, 51-56, 58-60]. This design took place after the microservices identification, and it refers to the definition of a domain model that reflects business knowledge isolated into each microservice. The third purpose shown in Fig. 3 was using Ubiquitous Language (UL) to cultivate a common language between domain experts and the development team to increase communication effectiveness [28, 34-37, 45, 59]. The fourth purpose was only identified in two primary studies [37, 38]. It uses a Subdomain pattern to split a broad business domain into more manageable partitions. Unlike microservices identification, the use of DDD for wide domain decomposition is about

reducing the complexity of the business domain through partitioning, where each part of the domain can be decomposed into several microservices.

(2) RQ-2: What is the evidence about the use of DDD for microservices-based systems development? The first one was software systems, and the second was models. Fig. 4 shows the systems developed by authors with DDD and Microservices Architecture. These systems were classified into two kinds based on the details mentioned by the authors about their development process and the context of the business domain problem.

As shown in Fig. 4, DDD was used to develop 36 microservices-based systems. Of these systems, 56,76% correspond to domains controlled and limited by authors to evaluate a proposal or explore the use of some DDD patterns and principles (Example systems) [27, 29-30, 40, 44, 55-59]. On the other hand, 43,24% of the mentioned systems correspond to real problems where it is necessary to satisfy the necessity of the clients (Industry systems) [26, 28, 34-38, 45-50, 52-53, 58, 60].

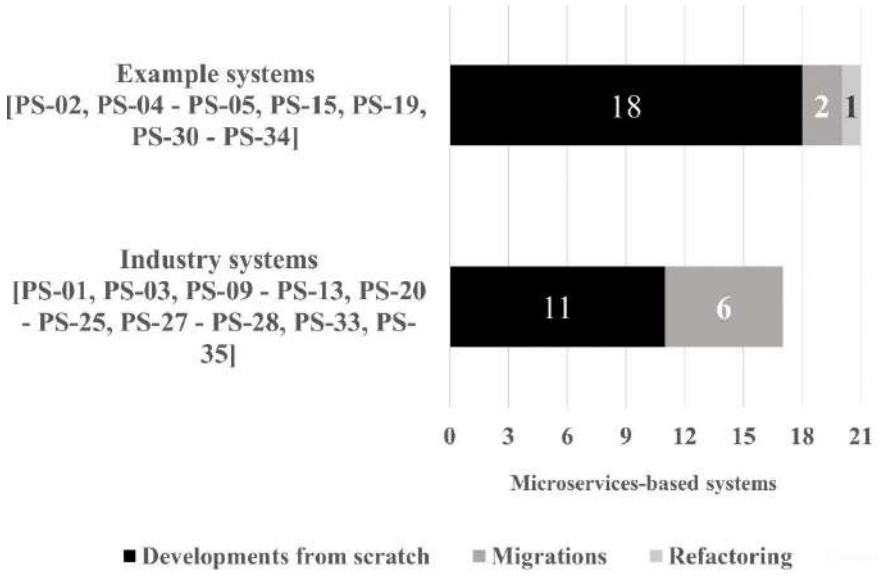


Fig. 4. Microservices-Based Systems developed with DDD.

On the other hand, a set of models involved in the microservices design process were identified. Some of them come from the DDD literature, but others were used to complement the design obtained with DDD, according to the authors of primary studies. In Fig. 5, a set of models is shown, classified by the type of system where the authors mentioned them.

Fig. 5 shows two DDD artifacts used during microservices design: the DDD Domain model and Context map. However, the DDD artifacts were not enough to deal all the specification aspects of microservices-based systems, reason why authors also used UML artifacts to complement the preliminary design obtained with DDD models.

Some models were created by following a notation proposed by authors of primary studies, such as the source model and sketching rough descriptions shown in Fig. 5. These artifacts do not seem to follow a clear standard or notation.

(3) RQ-3: What DDD patterns are used in the microservices-based systems development? A sum of 12 DDD patterns was used by the authors of primary studies in their microservices-based systems design process. These patterns are shown in Fig. 6.

Based on Fig. 6, strategic patterns were mentioned mainly in industry systems, while tactical patterns predominate in example systems. Fig. 6 shows BC as the DDD pattern most frequently mentioned in primary studies [27-30, 34-38, 40, 44-48, 52, 54, 60]. This pattern was treated as a microservice representation, and such as the definition by Evans [2], it delimits the scope of a model. The UL

pattern was used by authors of primary studies [28, 34-37, 45-46, 59] to increase the effectiveness of communication between domain experts and the development team, enabling a clear understanding of the problem. In addition, it is one of the patterns (together with Subdomain, ACL, and CS) that was only mentioned by authors who developed an industry system.

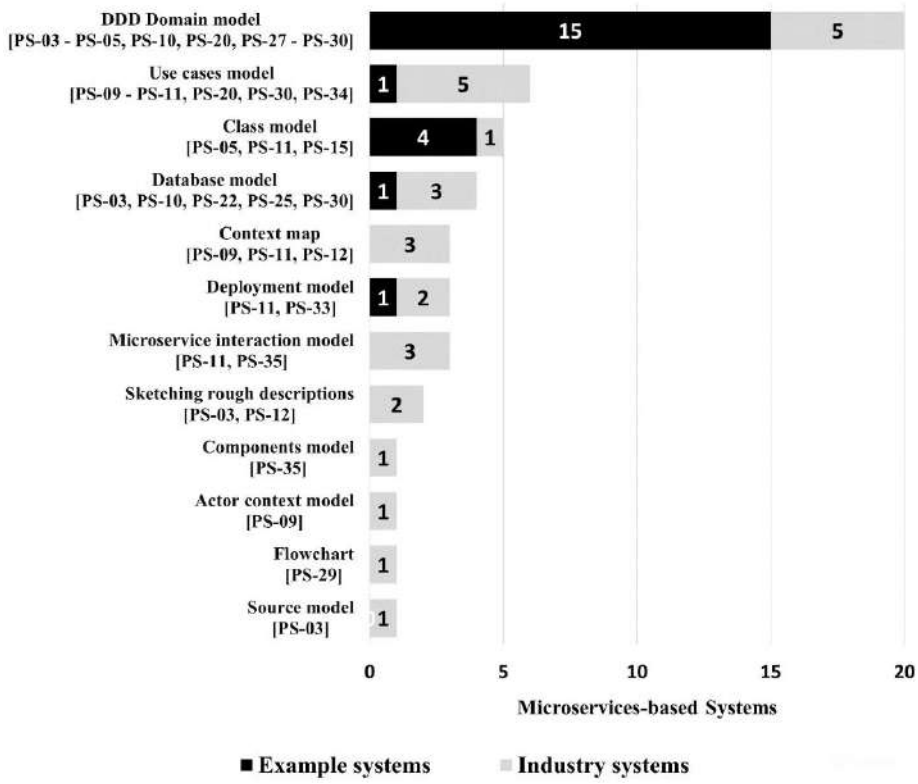


Fig. 5. Models used with DDD for microservices design.

Regarding tactical patterns, they were presented in a DDD domain model to obtain a domain-oriented microservices design. Entity was the tactical pattern most frequently mentioned by authors of primary studies [28-30, 35, 37, 45, 54, 60], and it was not always related to the Aggregate pattern. Unlike Entity, Aggregate is a pattern that requires using Entity and, optionally, other patterns such as Value object, Domain service, Repository, and others. Value object and Domain service were mentioned only as building blocks of the Aggregate pattern. Another pattern used with Aggregate was Repository, which was responsible for manipulating persistent data of an Aggregate through ACID transactions (Atomicity, Consistency, Isolation, and Durability). Event-Sourcing was a pattern mentioned during the DDD design [30], but no details were given about its usage in the microservices design.

(4) RQ-4: What technologies are used with DDD for microservices-based systems development? As reported by the authors, a set of technologies was identified in the microservices-based systems developments with DDD. Most of the technologies were used to implement microservices-based systems, and only three were reported as complements of the design with DDD [30, 34]. They are shown in Table 1.

(5) RQ-05: What techniques are used with DDD in the microservices-based systems development? The techniques of DDD used to complement DDD identified during the microservices-based systems development are shown in Fig. 7.

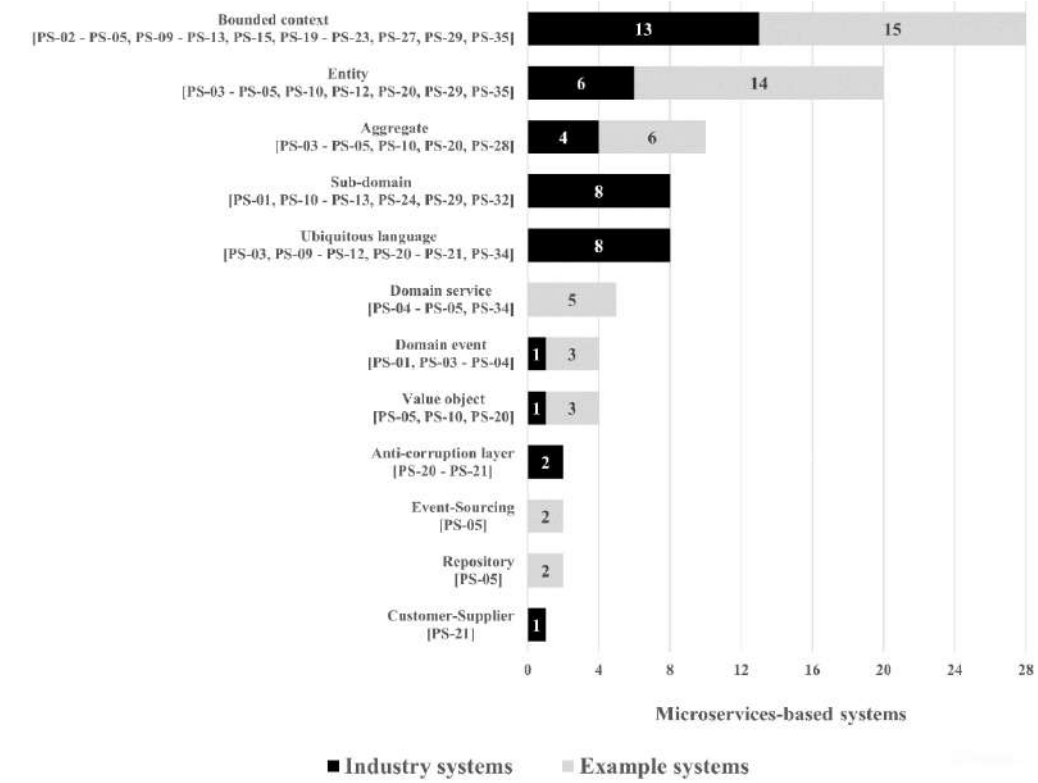


Fig. 6. Domain-Driven Design patterns used in microservices-based systems development.

Table 1. Technologies used with DDD for microservices-based systems development.

Technology name	Description	Primary studies	Official web site
Eclipse Papyrus	Design environment used by authors for code derivation from a DDD domain model made with a UML Profile	PS-05	https://eclipse.dev/papyrus/
ExplorViz	The 3D tool used by authors to identify coupling degrees between BCs (microservices).	PS-09	https://explorviz.dev/
Structure 101	Static code analysis tool used to scan a legacy monolithic project and obtain BC candidates.	PS-09	https://structure101.com/

As shown in Fig. 7, the authors used two kinds of techniques during the microservices design: elicitation techniques and DDD techniques. Context mapping was the DDD technique most frequently used by authors to model microservices candidates as BCs in a context map [34, 36-35, 46, 52]. Event-Storming [4] was a technique related to DDD, as mentioned in PS-01 [26], to identify subdomains, where each subdomain was considered a microservice. Although the authors of PS-01 [26] mentioned the work product obtained after Event-Storming execution (DDD subdomains), no details were mentioned about the procedure followed to perform Event-Storming. Regarding elicitation techniques, these were used together with UL [26, 28, 34, 36-37, 59, 60]. There are some strategies described in DDD literature to cultivate a UL, such as Domain storytelling, Knowledge crunching, and others. However, the authors of primary studies used interviews (mainly), brainstorming, focus groups, and questionnaires to extract the domain knowledge from the interaction with domain experts.

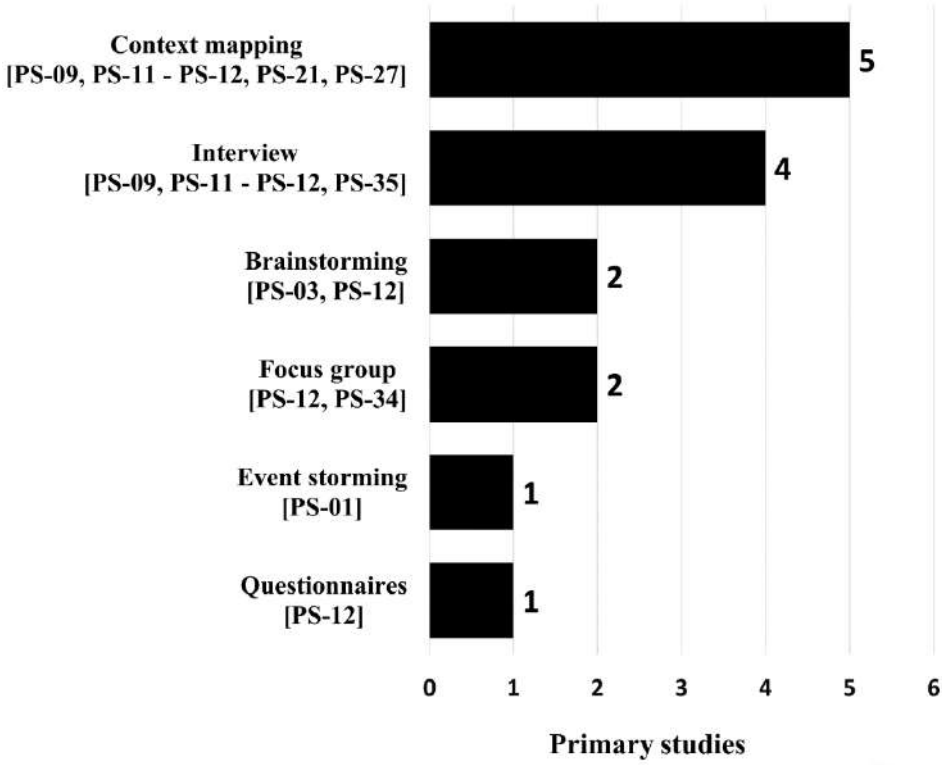


Fig. 7. Techniques used together with DDD.

(6) RQ-06: What challenges are mentioned by authors during the development of microservices-based systems? The authors mentioned 13 challenges when they used DDD in their development processes. These challenges are shown in Fig. 8.

The challenges identified were classified into six categories based on the problems that the authors described in primary studies. As shown in Fig. 10, the authors of primary studies mention two main difficulties. The first one is the procedure that is not defined [27, 30, 45]. It consists of the lack of a strict process to apply DDD techniques and patterns "correctly." The decision of what DDD pattern should be used and how depends on the business domain and the comprehension of a software engineer about the context of the problem. The second main challenge mentioned by authors of primary studies is related to the limitations mentioned by Evans [2]. When technical complexity predominates over the complexity of the business domain, DDD can complicate the solution [30, 42, 56]. The authors mentioned this because they do not recommend using DDD for developments where the most significant complexity is technical.

(7) RQ-7: What proposals exist for the development of microservices-based systems with DDD? Due to the incremental use of DDD in microservices-based systems development, some authors have proposed procedural guidelines to overcome the most frequent challenges of Microservice Architecture with the helplessness of DDD. These proposals are shown in Fig. 9, and they were classified according to the proposal type described by the authors of primary studies where they were extracted.

As shown in Fig. 9, 21 proposals were identified and classified into five categories. These categories come from the denomination authors use to refer to their proposals. For example, the authors named their proposals "Approaches" in six primary studies [35, 40, 43, 55-57]. In five primary studies [30, 33, 37, 39, 58], authors named their proposals as "Methodologies" and so on. Based on Fig. 9, it is also possible to see that 80,95% of proposals were evaluated by authors. In comparison, 19.05%

were not evaluated, postponing their evaluation to future works or delegating the evaluation for interested lectures.

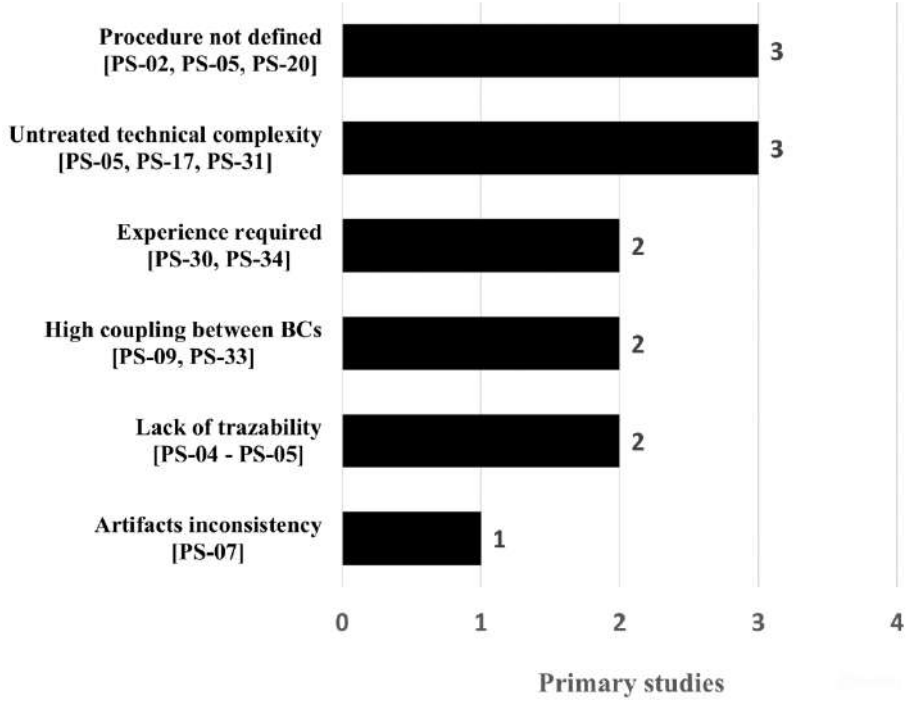


Fig. 8. Challenges faced with DDD in microservices-based systems development.

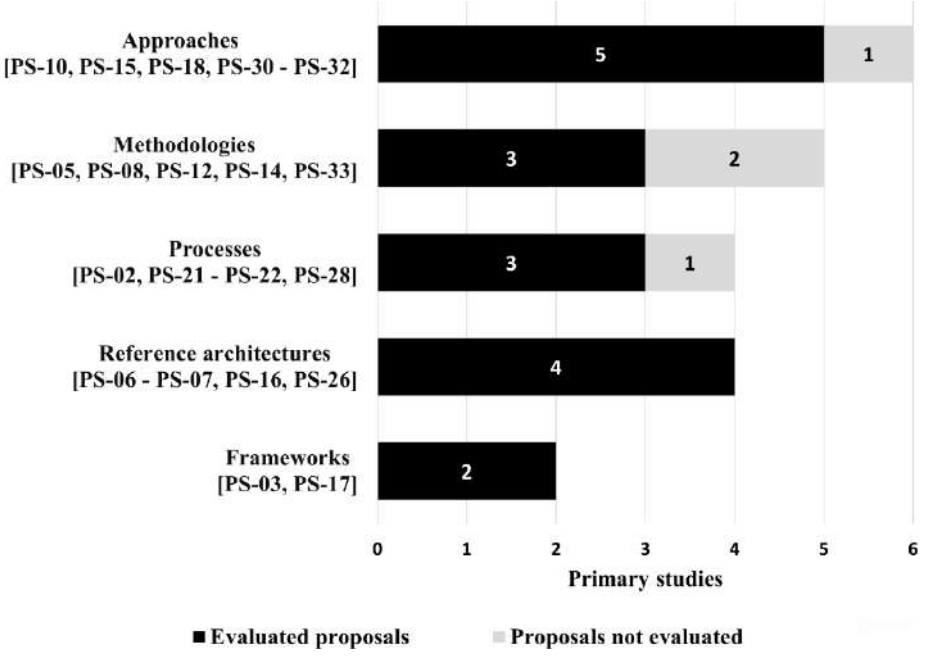


Fig. 9. Proposals for microservices design with DDD.

4.2 Thematic Synthesis Results

As a result of the systematic literature review, a familiarization phase recommended by Braun and Clark [24] was performed. However, a new data extraction was conducted based on RQs and thematic synthesis guidelines [23]. This data collection enabled us to identify meaning patterns among the data. This first level resulted in a set of code concepts seen as building blocks of themes. A sum of 32 codes were identified from the evidence. These codes were transformed into 11 themes that isolate the idea behind a group of codes. In the end, five higher-order themes were identified through theme grouping. Based on the thematic synthesis process, it is possible to describe a particular story of collected data, as mentioned by Braun and Clark [24]. This high-level overview synthesized with thematic synthesis can be seen in Fig. 10 in the high-order themes model.

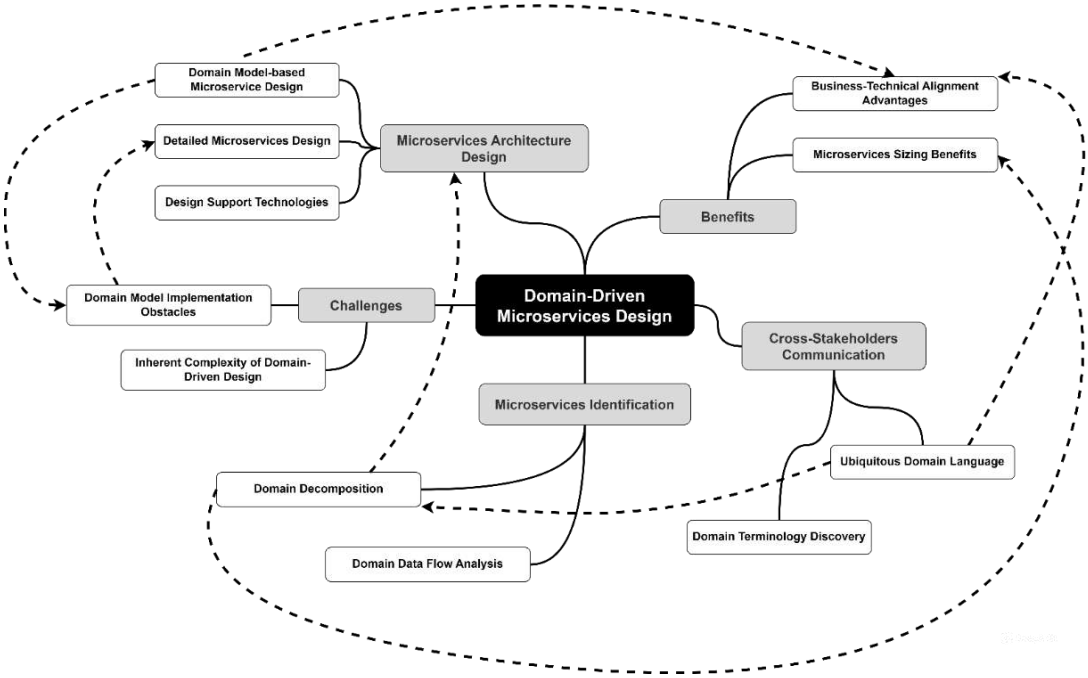


Fig. 10. Thematic Map of Domain-Driven Microservices Design.

(1) Cross-Stakeholders Communication. According to Vlad Khononov [4], the central idea of DDD is the communication. The domain knowledge shared between domain experts and developers should be clear and consistent. With this similar purpose, authors of primary studies familiarized themselves with the domain experts' jargon and used it to cultivate a UL free of technical details and ambiguous terms. This approach forms the basis of the theme "Ubiquitous Domain Language", which consists of using UL as a business domain glossary to enrich the domain knowledge exchange between stakeholders. This language is product of a distillation process, which is the idea behind the "Domain Terminology Discovery" theme, and it is related to the use of elicitation techniques mentioned in above sections.

(2) Microservice identification. Cultivating UL enables benefits related to effective communication, but another consequence of its usage is the identification of BCs. Each BC acts as a semantic boundary that delimits the meaning of the terms that conform to an UL. Through BCs, it is possible to decompose a business domain into semantic domain partitions that represent microservices. The theme "Domain Decomposition" encompasses all the activities and strategies (described in the above sections) used by authors of primary studies for business domain partitioning into BCs, subdomains, or Aggregates that represent microservices.

Another strategy was identified using DDD analysis techniques to identify clusters of domain concepts. The authors performed Event-storming to identify microservices candidates. This technique and the use of domain events reflect the "Domain Data Flow Analysis" theme, which involves the analysis of the closely related domain events that allowed authors to identify clusters treated as microservices.

(3) Microservices Architecture Design. As shown in the above sections, all tactical design patterns were translated into using the DDD domain model to design the business domain layer for each microservice identified with strategic design. The domain layer is a crucial part of the architecture of the microservices-based systems developed by authors, which is why the theme "Domain Model-Based Microservices Design" was defined. However, some other technical details were not specified with DDD artifacts.

Mainly in industry systems development, challenges related to the design specification of microservices were mentioned. This lack of technical specification for microservices was why other design patterns such as CQRS, Saga, Strangler Fig, and others were used together with standardized diagrams to describe details related to the implementation of microservices. These design resources used to refine the preliminary design obtained with DDD were defined as the "Detailed Microservices Design" theme. We also notice another design resource to refine the design obtained with DDD for each microservice. This is the use of technologies mentioned in primary studies PS-05 and PS-09. This action to complement the design of microservices was defined as the "Design Support Technologies" theme.

(4) Challenges. In answer to RQ6, challenges mentioned by authors of primary studies were extracted. As a result of thematic analysis, these challenges were classified into two themes that represent the two main difficulties faced by developers during microservices design with DDD. The theme named "Inherent Complexity of Domain-Driven Design" is related to the lack of guidelines, checkpoints, and a strict path to know if a developer is applying DDD correctly. Another challenge was defined as "Domain Model Implementation Obstacles", which comes from the problems faced by authors who tried to implement the DDD domain models. Some authors have made some proposals; however, there are no rules, guidelines, or strict specific ways to generate code from these DDD artifacts.

(5) Benefits. Just as the authors of primary studies have reported challenges in the use of DDD for microservices-based systems development, some authors mentioned the benefits obtained from the execution of some of DDD techniques and the use of its patterns. Some authors of primary studies mentioned benefits related to development complexity. In PS-04, PS-05, and PS-30, authors described the business domain complexity isolated into some DDD patterns such as BCs, Aggregates, or Subdomains. This isolation enabled them to tackle the most significant complexity of their microservices-based projects, the business domain logic. These benefits were grouped into the theme "Business-Technical Alignment Advantages".

Furthermore, other primary study authors mentioned benefits during the microservices size definition. Based on the decomposition process followed by authors, each microservice could sometimes be represented as a BC or an Aggregate. This decomposition proposed by DDD contributes to modifiability. These benefits related to the size of microservices were grouped in the theme of "Microservices Sizing Benefits".

5. Discussion

In this study, we successfully answered all the research questions by employing the research method conduction. Our efforts involved collecting and synthesizing a wealth of knowledge on the practical use of Domain-Driven Design (DDD) in developing microservices-based systems.

Analyzing the demographic results of the study yielded interesting findings, particularly an increased interest in the adoption of DDD in Microservices Architecture (MSA). It is worth noting that a gap exists between theoretical understanding and practical implementation of certain patterns,

such as Subdomain or BC. Consequently, through this research, we have provided evidence-based knowledge on these patterns and their application. Our findings complement the grounded theory study published by Singjai et al. [14] and the systematic review conducted by Schmidt et al. [15], offering valuable insights into the practical use of DDD in microservices system design.

The utilization of Domain-Driven Design (DDD) has emerged as a vital component in the domain analysis phase of microservices-based systems development within the industry. Strategic design, in particular, plays a crucial role in establishing a shared understanding among stakeholders, enabling authors to express ideas unambiguously. Conversely, developers have primarily utilized tactical design to tackle controlled domain problems and serve specific purposes. Additionally, existing literature indicates that DDD has been employed to decompose business domains into microservices candidates in the analysis process. However, it is important to note that the BC pattern is not the only one utilized or emphasized in the literature. Using UL for stakeholder interaction is a common practice in complex domains where developers may not be familiar with the domain. On the other hand, applying Tactical design in industrial projects has been less frequent. Thus, certain DDD patterns, such as Domain Event, Event-Sourcing, and Domain Services, remain underutilized in real-world contexts.

6. Threats to validity and limitations

In the literature reviews, Kitchenham and other authors [19, 22-23] emphasized the importance of reliability. This aspect was carefully considered throughout the research process, from manual search to data synthesis using Cruzes and Dyba's proposal. We implemented a series of mitigation measures to minimize potential biases at various stages of the research.

To ensure the selection of relevant papers was unbiased, we utilized a manual search approach and established inclusion and exclusion criteria based on the Quasi-Gold Standard. These criteria helped us avoid solely relying on one search engine's studies. Once we identified primary studies, we further augmented our research by employing a snowballing technique. This process helped to minimize the possibility of overlooking any relevant studies.

Once the selection process was complete, the chosen primary studies underwent a rigorous evaluation by the authors of this study to ensure their relevance to at least one RQ. Also, the authors continuously reviewed and revised their work during the data extraction process to maintain accuracy. Review questions were developed and regularly evaluated to avoid omissions and confirm that no crucial data had been missed. The same meticulous approach was applied when defining themes and subthemes, with each code being meticulously linked to specific text segments and the themes closely tied to these codes. In the same way, the names assigned to the codes, themes, and higher-order themes were determined through collaborative revisions among the authors of this study.

7. Conclusion

In this study, we adopted the systematic literature review method proposed by Kitchenham [19] to examine the utilization of DDD in developing a microservices-based system. We formulated seven research questions (RQs) to guide our research process and ensure focused research. Our selection process involved both manual and automatic searches to identify relevant studies. Through this process, we identified 31 primary studies. We also employed snowballing techniques to enhance our selection, which led us to four additional studies. We then conducted a preliminary synthesis to familiarize ourselves with the primary studies and address the RQs, mainly focusing on the application of DDD in the development of microservices-based systems. To gather the necessary data, we performed an extraction process. To provide a comprehensive analysis, we further conducted a thematic synthesis utilizing the method proposed by Cruzes and Dyba. To complement this approach, we also incorporated recommendations from the Braun and Clark proposal, ensuring a robust analysis of the collected data.

Throughout our analysis, we have identified specific details regarding the application of DDD that contribute to enhancing effective knowledge sharing between developers and domain experts. These details primarily revolve around the integration of UL with DDD and the utilization of various elicitation techniques. Interestingly, these aspects have not been extensively addressed in related studies, thereby providing fresh insights into the broader scope of DDD beyond its traditional utilization for system decomposition.

Among the different uses we discovered, the most frequently reported one involves decomposing a business domain or legacy system into microservices. However, our analysis captured new and pertinent details about using strategic patterns to define the business scope of microservices, as well as variations and adaptations.

Most authors in the primary studies highlighted the successful implementation of microservices, explicitly noting the absence of coupling issues between microservices. Some authors went so far as to underscore DDD's potential for achieving an optimal scope of microservices based on business capabilities. While the remaining authors did not mention any problems in their DDD-driven microservices systems, they did not specifically address certain characteristic aspects of DDD within the context of MSA.

Despite the overall positive outcomes reported, some challenges persist in the practical application of DDD. These challenges primarily stem from the perceived complexity of implementing DDD, which can be particularly daunting for developers without prior experience analyzing and designing intricate business domains. Additionally, there is an opportunity for future work in refining the implementation of DDD artifacts, such as the domain model, to further enhance its effectiveness and efficiency in microservices development. Finally, we envision future work focused on delving into the creation of DDD patterns that allow the development of code that effectively represents the underlying business logic, with minimal dependencies on specific programming languages based on Object-Oriented Programming.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

[1]. J. Sangabriel-Alarcón, J. O. Ocharán-Hernández, K. Cortés-Verdín, and X. Limón, “Domain-Driven Design for Microservices Architecture Systems Development: A Systematic Mapping Study,” in 2023 11th International Conference in Software Engineering Research and Innovation (CONISOFT), 2023, pp. 25–34.

[2]. E. Evans, “Domain-driven design: tackling complexity in the heart of software,” p. 529, 2004.

[3]. E. Evans, “Domain-Driven Design Reference: Definitions and Pattern Summaries”, 2014.

[4]. V. Khononov and J. Lerman, “Learning domain-driven design: aligning software architecture and business strategy”, p. 312, 2021.

[5]. V. Vernon, “Implementing Domain-Driven Design”, 2013.

[6]. V. Velepucha and P. Flores, “Monoliths to microservices-Migration Problems and Challenges: A SMS,” Proceedings - 2021 2nd International Conference on Information Systems and Software Technologies, ICI2ST 2021, pp. 135–142, Mar. 2021, doi: 10.1109/ICI2ST51859.2021.00027.

[7]. G. Liu, B. Huang, Z. Liang, M. Qin, H. Zhou, and Z. Li, “Microservices: Architecture, container, and challenges,” Proceedings - Companion of the 2020 IEEE 20th International Conference on Software Quality, Reliability, and Security, QRS-C 2020, pp. 629–635, Dec. 2020, doi: 10.1109/QRS-C51114.2020.00107.

[8]. R. Mubashir, J. Ahmed, F. Khakwani, and T. Rana, Microservices Architecture: Challenges and Proposed Conceptual Design. 2019.

[9]. S. Salii, J. Ajdari, and X. Zenuni, “Migrating to a microservice architecture: benefits and challenges,” 2023.

[10]. S. Newman, “Building microservices: Designing fine-grained systems (second edition),” pp. 1–10, 2021.

[11]. F. Rademacher, J. Sorgalla, and S. Sachweh, “Challenges of domain-driven microservice design: A model-driven perspective,” IEEE Softw, vol. 35, no. 3, pp. 36–43, May 2018, doi: 10.1109/MS.2018.2141028.
[12]. M. Tello-Rodríguez, J. O. Ocharán-Hernández, J. C. Pérez-Arriaga, X. Limón, and Á. J. Sánchez-García, “A Design Guide for Usable Web APIs,” Programming and Computer Software, vol. 46, no. 8, pp. 584–593, 2020, doi: 10.1134/S0361768820080241.
[13]. B. Jin, S. Sahni, and A. Shevat, “Designing Web APIs,” 2018.
[14]. A. Singjai, U. Zdun, and O. Zimmermann, “Practitioner Views on the Interrelation of Microservice APIs and Domain-Driven Design: A Grey Literature Study Based on Grounded Theory,” in Proceedings - IEEE 18th International Conference on Software Architecture, ICSA 2021, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 25–35. doi: 10.1109/ICSA51549.2021.00011.
[15]. R. A. Schmidt and M. Thiry, “Microservices identification strategies: A review focused on Model-Driven Engineering and Domain Driven Design approaches,” in 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), IEEE, 2020. Accessed: Nov. 13, 2022. [Online]. Available: https://ieeexplore-ieee-org.ezproxy.uv.mx/document/9141150/.
[16]. A. Macias, E. Navarro, C. Cuesta, and U. Zdun, “Architecting Digital Twins Using a Domain-Driven Design-Based Approach,” International Conference on Software Architecture (ICSA), no. 62, pp. 183–209, 2023, doi: 10.13039/501100011033.
[17]. C. Praschl, S. Bauernfeind, C. Leitner, and G. A. Zwettler, “Domain-Driven Design as Model Contract in Full-Stack Development,” in International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICECCME57830.2023.10252654.
[18]. F. Rademacher, S. Sachweh, and A. Zündorf, “Towards a UML profile for domain-driven design of microservice architectures,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10729 LNCS, pp. 230–245, 2018, doi: 10.1007/978-3-319-74781-1_17/COVER.
[19]. B. A. Kitchenham, D. Budgen, and P. Brereton, “Evidence-Based Software Engineering and Systematic Reviews,” 2015.
[20]. H. Zhang, M. A. Babar, and P. Tell, “Identifying relevant studies in software engineering,” Inf Softw Technol, vol. 53, no. 6, pp. 625–637, Jun. 2011, doi: 10.1016/j.infsof.2010.12.010.
[21]. C. Wohlin, “Guidelines for snowballing in systematic literature studies and a replication in software engineering,” in Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE ’14, New York, New York, USA: ACM Press, 2014, pp. 1–10. doi: 10.1145/2601248.2601268.
[22]. J. Popay et al., “Guidance on the conduct of narrative synthesis in systematic Reviews. A Product from the ESRC Methods Programme. Version 1,” undefined, 2006, doi: 10.13140/2.1.1018.4643.
[23]. D. S. Cruzes and T. Dybá, “Recommended steps for thematic synthesis in software engineering,” in International Symposium on Empirical Software Engineering and Measurement, IEEE Computer Society, 2011, pp. 275–284. doi: 10.1109/esem.2011.36.
[24]. V. Clarke and V. Braun, “Thematic analysis: A practical guide,” London: SAGE, 2021, Accessed: Oct. 30, 2023. [Online]. Available: https://uk.sagepub.com/en-gb/eur/thematic-analysis/book248481#description.
[25]. J. Sangabriel-Alarcón, J. O. Ocharán-Hernández, X. Limón, and K. Cortés-Verdín, “Domain-Driven Design in Microservices-Based Systems Development: A Systematic Literature Review and Thematic Analysis [Dataset].” [Online]. Available: https://zenodo.org/records/13294975.
[26]. PS-01. G.-C. Pan, P. Liu, and J.-J. Wu, “A Cloud-Native Online Judge System,” in 2022 IEEE COMPSAC, 2022, doi: 10.1109/COMPSAC54236.2022.00204.
[27]. PS-02. N. Ivanov and A. Tasheva, “A Hot Decomposition Procedure: Operational Monolith System to Microservices,” in 2021 IEEE ICAI, 2021, doi: 10.1109/ICAI52893.2021.9639494.
[28]. PS-03. M. I. Joselyne, G. Bajpai, and F. Nzanywayingoma, “A Systematic Framework of Application Modernization to Microservice-based Architecture,” in 2021 IEEE ICEET, 2021, doi: 10.1109/ICEET53442.2021.9659783.
[29]. PS-04. A. Singjai and U. Zdun, “Conformance Assessment of Architectural Design Decisions on the Mapping of Domain Model Elements to APIs and API Endpoints,” in 2022 IEEE ICSA-C, 2022, doi: 10.1109/ICSA-C54293.2022.00058.
[30]. PS-05. F. Rademacher, S. Sachweh, and A. Zündorf, “Deriving Microservice Code from Underspecified Domain Models Using DevOps-Enabled Modeling Languages and Model Transformations,” in 2020 IEEE SEAA, 2020, doi: 10.1109/SEAA51224.2020.00047.

[31]. PS-06. A. Steffens, H. Lichter, and J. S. Döring, "Designing a Next-Generation Continuous Software Delivery System: Concepts and Architecture," in 2018 ACM Conference on Software Engineering and Applications, 2018, doi: 10.1145/3194760.3194768.
[32]. PS-07. Y. Ding et al., "Enterprise Service Application Architecture Based on Domain Driven Model Design," in 2020 IEEE ITCA, 2020, doi: 10.1109/ITCA52113.2020.00167.
[33]. PS-08. P. Ray and P. Pal, "Extending the SEMAT Kernel for the Practice of Designing and Implementing Microservice-Based Applications using Domain Driven Design," in 2020 IEEE CSEET, 2020, doi: 10.1109/CSEET49119.2020.9206200.
[34]. PS-09. A. Krause et al., "Microservice Decomposition via Static and Dynamic Analysis of the Monolith," in 2020 IEEE ICSA-C, 2020, doi: 10.1109/ICSA-C50368.2020.00011.
[35]. PS-10. C.-Y. Li, S.-P. Tseng, and T.-W. Lu, "Microservice Migration Using Strangler Fig Pattern: A Case Study on the Green Button System," in 2020 IEEE ICS, 2020, doi: 10.1109/ICS51289.2020.00107.
[36]. PS-11. A. Rahmatulloh et al., "Microservices-Based IoT Monitoring Application with a Domain-Driven Design Approach," in 2021 IEEE ICADEIS, 2021, doi: 10.1109/ICADEIS52521.2021.9701966.
[37]. PS-12. M. I. Josélyne et al., "Partitioning Microservices: A Domain Engineering Approach," in 2018 ACM Conference on Software Engineering and Applications, 2018, doi: 10.1145/3195528.3195535.
[38]. PS-13. M. Pham and D. B. Hoang, "SDN Applications - The Intent-Based Northbound Interface Realization for Extended Applications," in 2016 IEEE NetSoft, 2016, doi: 10.1109/NETSOFT.2016.7502469.
[39]. PS-14. E. Cabrera et al., "Towards a Methodology for Creating Internet of Things (IoT) Applications Based on Microservices," in 2020 IEEE SCC, 2020, doi: 10.1109/SCC49832.2020.00072.
[40]. PS-15. R. Petrasch, "Model-Based Engineering for Microservice Architectures Using Enterprise Integration Patterns for Inter-Service Communication," in 2017 IEEE JCSSE, 2017, doi: 10.1109/JCSSE.2017.8025912.
[41]. PS-16. J. Dobaj et al., "A Microservice Architecture for the Industrial Internet-Of-Things," in 2018 ACM Conference on Software Engineering and Applications, 2018, doi: 10.1145/3282308.3282320.
[42]. PS-17. S. Braun, A. Bieniusa, and F. Elberzhager, "Advanced Domain-Driven Design for Consistency in Distributed Data-Intensive Systems," in 2021 ACM Conference on Software Engineering and Applications, 2021, doi: 10.1145/3447865.3457969.
[43]. PS-18. M. Khemaja, "Domain Driven Design and Provision of Micro-Services to Build Emerging Learning Systems," in 2016 ACM Conference on Software Engineering and Applications, 2016, doi: 10.1145/3012430.3012643.
[44]. PS-19. Z. Li, "Using Public and Free Platform-as-a-Service (PaaS) Based Lightweight Projects for Software Architecture Education," in 2020 ACM Conference on Software Engineering and Applications, 2020, doi: 10.1145/3377814.3381704.
[45]. PS-20. P. Oukes et al., "Domain-Driven Design Applied to Land Administration System Development: Lessons from the Netherlands," in Land Use Policy, vol. 105, 2021, doi: 10.1016/j.landusepol.2021.105379.
[46]. PS-21. C. E. da Silva, Y. de Lima Justino, and E. Adachi, "SPReaD: Service-Oriented Process for Reengineering and DevOps," in Software: Practice and Experience, 2022, doi: 10.1007/s11761-021-00329-x.
[47]. PS-22. C.-Y. Fan and S.-P. Ma, "Migrating Monolithic Mobile Application to Microservice Architecture: An Experiment Report," in 2017 IEEE AIMS, 2017, doi: 10.1109/AIMS.2017.23.
[48]. PS-23. A. Krylovskiy, M. Jahn, and E. Patti, "Designing a Smart City Internet of Things Platform with Microservice Architecture," in 2015 IEEE FiCloud, 2015, doi: 10.1109/FiCloud.2015.55.
[49]. PS-24. K. Zhang et al., "Design of Domain-Driven Microservices-Based Software Talent Evaluation and Recommendation System," in 2022 IEEE ICEKIM, 2022, doi: 10.1109/ICEKIM55072.2022.00076.
[50]. PS-25. Q. Li, W. Sun, and R. Ma, "Sharing Platform of Digital Specimen of Wood Canker Based on WebGIS in Xinjiang Province: Architecture, Design and Implementation," in 2022 IEEE CIPAE, 2022, doi: 10.1109/CIPAE55637.2022.00029.
[51]. PS-26. T. Raffin et al., "A Reference Architecture for the Operationalization of Machine Learning Models in Manufacturing," in Procedia CIRP, vol. 2022, 2022, doi: 10.1016/j.procir.2022.10.062.
[52]. PS-27. C. Batista et al., "Towards a Multi-Tenant Microservice Architecture: An Industrial Experience," in 2022 IEEE COMPSAC, 2022, doi: 10.1109/COMPSAC54236.2022.00100.
[53]. PS-28. C. Praschl et al., "Domain-Driven Design as Model Contract in Full-Stack Development," in 2023 IEEE ICECCME, 2023, doi: 10.1109/ICECCME57830.2023.10252654.

[54]. PS-29. N. Legowo et al., “Designing Service Oriented Architecture Model in Sehatin Application with a Domain-Driven Design Approach,” in 2023 IEEE ICIMTech, 2023, doi: 10.1109/ICIMTech59029.2023.10278057.
[55]. PS-30. A. Macías et al., “Architecting Digital Twins Using a Domain-Driven Design-Based Approach,” in 2023, doi: 10.13039/501100011033.
[56]. PS-31. I. V. P. and V. P. H., “An Approach to Clean Architecture for Microservices Using Python,” in 2023 IEEE CSITSS, 2023, doi: 10.1109/CSITSS60515.2023.10334229.
[57]. PS-32. M. Saidi, A. Tissaoui, and S. Faiz, “A DDD Approach Towards Automatic Migration To Microservices,” in 2023 IEEE IC_ASET, 2023, doi: 10.1109/IC_ASET58101.2023.10150522.
[58]. PS-33. M. Camilli et al., “Actor-Driven Decomposition of Microservices through Multi-Level Scalability Assessment,” in 2023, doi: 10.1145/3583563.
[59]. PS-34. O. Özkan, Ö. Babur, and M. van den Brand, “Refactoring with Domain-Driven Design in an Industrial Context: An Action Research Report,” in Software: Practice and Experience, 2023, doi: 10.1007/s10664-023-10310-1.
[60]. PS-35. E. T. Nordli et al., “Migrating Monoliths to Cloud-Native Microservices for Customizable SaaS,” in Information and Software Technology, vol. 2023, 2023, doi: 10.1016/j.infsof.2023.107230.

Информация об авторах / Information about authors

Хосуэ САНГАБРИЭЛЬ-АЛАРКОН – инженер-программист, разработчик программного обеспечения Университете Веракруса (Мексика). Сфера научных интересов: архитектура программного обеспечения, проектирование программного обеспечения, инженерия требований, моделирование данных.

Josué SANGABRIEL-ALARCÓN – Software Engineer. Software Developer at Universidad Veracruzana, Mexico (University of Veracruz). Research interests: software architecture, software design, requirements engineering, data modeling.

Хорхе Октавио ОЧАРАН-ЭРНАНДЕС имеет степень PhD по программированию, доцент факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных интересов: архитектура программного обеспечения, инженерия требований, программная инженерия, разработка прикладных интерфейсов.

Jorge Octavio OCHARÁN-HERNÁNDEZ – PhD in Computer Science, Associate Professor at the Facultad de Estadística e Informática, Universidad Veracruzana, Mexico (School of Statistics and Informatics, University of Veracruz) since 2017. Research interests: software architecture, requirements engineering, software engineering, API design.

Ксавьер ЛИМОН имеет степень PhD по искусственному интеллекту, доцент факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных интересов: интеллектуальный анализ данных, мультиагентные и распределенные системы, архитектура программного обеспечения.

Xavier LIMÓN – PhD in Artificial Intelligence, Associate Professor at Facultad de Estadística e Informática, Universidad Veracruzana, Mexico (School of Statistics and Informatics, University of Veracruz). Research interests: data mining, multiagent systems, distributed systems, software architecture.

М. Карен КОРТЕС-ВЕРДИН имеет степень PhD по искусственному интеллекту, профессор факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных интересов: программные продуктовые линии, архитектуры программного обеспечения, аспектно-ориентированное программирование, разработка программного обеспечения, ориентированного на решение конкретных задач, процессы программирования, качество программного обеспечения, моделирование программ.

María Karen CORTÉS-VERDÍN – PhD in Computer Science, Professor at School of Statistics and Informatics, Universidad Veracruzana, Mexico. Research interests: software product lines, software architectures, aspect-oriented software development, concern-oriented software development, software process, software quality, software modeling.

DOI: 10.15514/ISPRAS-2024-36(6)-4

A Decade of Advancements in Program Synthesis from Natural Language: A Systematic Literature Review

[1] *R. Ramirez-Rueda, ORCID: 0009-0004-2084-7505 <zS20000354@estudiantes.uv.mx>*
[1] *E. Benitez-Guerrero, ORCID: 0000-0001-5844-4198 <edbenitez@uv.mx>*
[1] *C. Mezura-Godoy, ORCID: 0000-0002-5386-107X <cmezura@uv.mx>*
[2] *E. Barcenas, ORCID: 0000-0002-1523-1579 <barcenas@fi-b.unam.mx>*

[1] *Facultad de Estadistica e Informatica, Universidad Veracruzana, Xalapa, Mexico.*
[2] *Universidad Nacional Autonoma de Mexico Ciudad de Mexico, Mexico.*

Abstract. Program Synthesis is the process of automatically generating software from a requirement specification. This paper presents a systematic literature review focused on program synthesis from specifications expressed in natural language. The research problem centers on the complexity of automatically generating accurate and robust code from high-level, ambiguous natural language descriptions – a barrier that limits the broader adoption of automatic code generation in software development. To address this issue, the study systematically examines research published between 2014 and 2024, focusing on works that explore various approaches to program synthesis from natural language inputs. The review follows a rigorous methodology, incorporating search strings tailored to capture relevant studies from five major data sources: IEEE, ACM, Springer, Elsevier, and MDPI. The selection process applied strict inclusion and exclusion criteria, resulting in a final set of 20 high-quality studies. The findings reveal significant advancements in the field, particularly in the integration of large language models (LLMs) with program synthesis techniques. The review also highlights the challenges and concludes by outlining key trends and proposing future research directions aimed at overcoming these challenges and expanding the applicability of program synthesis across various domains.

Keywords: program synthesis; program generation; natural language processing.

For citation: Ramirez-Rueda R., Benitez-Guerrero E., Mezura-Godoy C., Barcenas E. A. Decade of Advancements in Program Synthesis from Natural Language: A Systematic Literature Review. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 59-82. DOI: 10.15514/ISPRAS-2024-36(6)-4.

Acknowledgments. This work was partially developed under the support of the National Council of Humanities, Science and Technology (CONAHCYT-Mexico) in the scope of the project “Infraestructura para Agilizar el Desarrollo de Sistemas Centrados en el Usuario” (Catedras, Ref. 3053). In addition, the authors thank CONAHCYT for the doctoral scholarship granted to the first author. We also thank the Universidad Veracruzana for the support in the development of this research.

Десятилетие достижений в синтезе программ по спецификациям на естественном языке: систематический обзор литературы

[1] *Р. Рамирес-Руэда, ORCID: 0009-0004-2084-7505 <zS20000354@estudiantes.uv.mx>*
[1] *Э. Бенитес-Гуэрреро, ORCID: 0000-0001-5844-4198 <edbenitez@uv.mx>*
[1] *К. Мезура-Годой, ORCID: 0000-0002-5386-107X <cmezura@uv.mx>*
[2] *Э. Барсенас, ORCID: 0000-0002-1523-1579 <barcenas@fi-b.unam.mx>*

[1] *Факультет статистики и информатики Университета Веракруса, Халапа, Мексика.*
[2] *Национальный автономный университет города Мехико, Мексика.*

Аннотация. Программный синтез – это процесс автоматического создания программного обеспечения на основе спецификации требований. В этой статье представлен систематический обзор литературы, посвященный синтезу программ из спецификаций, выраженных на естественном языке. Исследуемая проблематика заключается в сложности автоматического создания точного и надежного кода из высокоуровневых, неоднозначных описаний на естественном языке – барьер, который ограничивает более широкое использование средств автоматизации при разработке программного обеспечения. Для исследования этой проблемы авторы систематически изучали работы, опубликованные в период с 2014 по 2024 год, делая акцент на работы, в которых рассматриваются различные подходы к синтезу программ на основе данных на естественном языке. Обзор следует строгой методологии, включающей поисковые строки, адаптированные для сбора соответствующих исследований из пяти основных источников данных: IEEE, ACM, Springer, Elsevier и MDPI. В процессе отбора применялись строгие критерии включения и исключения, что привело к окончательному набору из 20 высококачественных исследований. Результаты показывают значительные достижения в этой области, особенно в интеграции больших языковых моделей (LLM) с методами синтеза программ. Обзор также освещает проблемы и завершается изложением ключевых тенденций и предложением будущих направлений исследований, нацеленных на преодоление этих проблем и расширение применимости синтеза программ в различных областях.

Ключевые слова: синтез программ; генерация программ; обработка естественного языка.

Для цитирования: Рамирес-Руэда Р., Бенитес-Гуэрреро Э., Мезура-Годой К., Барсенас Э. Десятилетие достижений в синтезе программ по спецификациям на естественном языке: систематический обзор литературы. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 59–82 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–4.

Благодарности. Эта работа была частично выполнена при поддержке Национального совета по гуманитарным наукам, науке и технике (CONAHCYT-Мексика) в рамках проекта “Infrastructura para Agilizar el Desarrollo de Sistemas Centrados en el Usuario” (Catedras, Ref. 3053). Кроме того, авторы благодарят совет CONAHCYT за докторскую стипендию, предоставленную первому автору. Мы также благодарим Университет штата Веракрус за поддержку в проведении этого исследования**.**

1. Introduction

The development of a software system encompasses a detailed life cycle that includes stages such as requirement specification, design, prototyping, and testing [1]. Traditionally, this process is slow and susceptible to errors. To enhance efficiency and reduce errors, employing models at various abstraction levels, along with their mappings, can facilitate the automatic generation of code from high-level descriptions. These models, which capture the behavior and structure of the system, can be developed manually or derived from requirement specifications. Despite the benefits, the process of generating code from models necessitates establishing the models themselves, defining rules for mapping elements between models, and creating rules to generate code in the target programming language. These tasks require expert knowledge and sophisticated tools.

An alternative method is program synthesis [2], which involves generating software automatically from a requirement specification. Relying on artificial intelligence and formal methods, this approach aims to produce correct programs by formulating rules that map input specifications

directly to programs, thus accelerating development. However, it is critical to acknowledge that software developed through this approach may be more prone to errors and could lack robustness.

In the realm of specifying systems, one commonly uses expressions in predicate logic, necessitating specialized expertise. To make this approach accessible not only to experts but also to end-users, specifications should ideally be articulated in a more intuitive form, such as natural language.

Our research takes as reference the work proposed by [3] and [4], which extensively review code generation with natural language, although they examine approaches that automatically generate source code from a description In natural language, we want to emphasize the areas of application, as well as to make known the types of inputs and outputs that are necessary to generate automatic code from natural language and finally analyze future trends.

To ensure the relevance of our study in this rapidly evolving domain, we consider recent advancements in natural language processing and artificial intelligence, particularly as they pertain to program synthesis. This includes the exploration of models like GPT-4 and other advanced transformer architectures to understand how they can be adapted for interpreting natural language program specifications. Moreover, we address the current challenges, such as achieving precision in interpreting complex requirements and the implications of automating code generation.

It is important to note that this work extends the paper "Program Synthesis and Natural Language Processing: A Systematic Literature Review," presented at the International Conference on Research and Innovation in Software Engineering (CONISOFT 2023). In this updated study, we expand the analysis by covering an additional five years and incorporating a new digital library (MDPI), thereby covering the last decade. Our objective is to analyze publications, identify emerging trends, and highlight opportunities for future research that were not addressed in the previous work. We selected twenty articles from major databases, including IEEE, ACM, Springer, Elsevier, and MDPI.

These studies investigate various methods of program synthesis, ranging from rule-based approaches, which employ explicit translation rules from natural language to code, to more advanced techniques that learn these rules from input-output pairs, integrating generative artificial intelligence models.

The paper is structured as follows: We begin with background information on the relevant research areas of program synthesis, natural language processing, and generative models. Next, we detail the methodology employed for the SLR, followed by a discussion of the findings. The paper concludes with a summary of the research outcomes.

2. Background

2.1 Program synthesis

In this section we discuss Program synthesis is an intriguing research domain focused on the automatic generation of programs from detailed specifications. This field is particularly valuable for creating small, complex programs that are verifiable and correct based on comprehensive specifications.

The domain is characterized by three critical dimensions [5]: the types of constraints that express user intentions, the operational search space, and the search techniques employed. User intentions can be depicted through various forms such as logical relations between inputs and outputs, demonstrations, natural language, input-output examples, or inefficient or related programs. The search space may be confined by potential program types, computational models like context-free grammars, or logical frameworks. Search techniques employed include exhaustive searches, version space exploration, machine learning, and logical reasoning.

Program synthesis can further be classified into methods such as deductive synthesis from full specifications [6], which generates programs based on probabilistic selection mechanisms. The viability of a program is determined by its alignment with specified criteria derived from its specifications. Despite its effectiveness, generating these detailed specifications is a considerable

challenge and verifying them is computationally intensive.

Alternately, inductive synthesis starts with incomplete problem descriptions, which may include test cases, specified desired and undesired behaviors, input-output examples, or execution traces for particular inputs [7]. While this approach ensures correctness by construction, the creation of extensive programs remains a significant computational challenge, often requiring more effort to define a complete and correct specification than to write the program itself. Fig. 1 shows the possible program synthesis approaches.

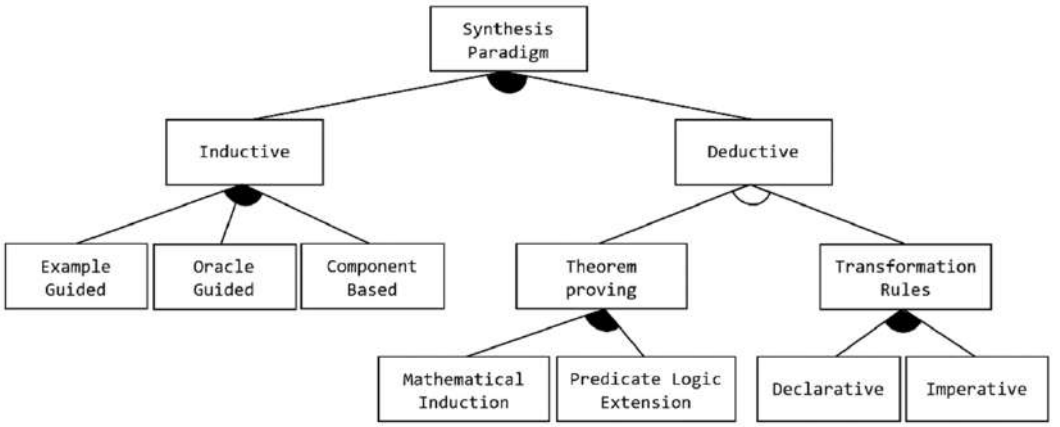


Fig. 1. Program Synthesis Paradigms [8].

Conversely, the integration of generative artificial intelligence is reshaping the software industry, not only by advancing techniques and tools but also by democratizing software development [9]. A significant obstacle in traditional program synthesis has been the requirement for complete specifications. However, modern advancements in software development are transforming this challenge by employing models capable of interpreting natural language descriptions to synthesize code across various programming languages. This transformation greatly simplifies the synthesis process and broadens access to those without specialized expertise.

A prime example is the use of large language models (LLMs), which empower non-programmers to create applications through intuitive natural language interfaces. This capability could herald a major shift in information technology education and training, with a greater emphasis on design and project management skills rather than on pure coding [10].

Unlike traditional expert systems that merely analyze or interact with existing data, program synthesis harnesses vast data sets and complex architectures to generate new and varied content. By leveraging continued advances in computing power, this approach employs deep neural networks, transformers, generative adversarial networks, and autoencoders to capture the complexity of data and effectively model high-dimensional probabilistic distributions across both specific and general domains [11].

Furthermore, by incorporating techniques that map the latent semantic space of language or images to multimedia representations in text, audio, or video, generative models can convert any type of input into a variety of output formats [12] [13]. This versatility makes generative models invaluable in numerous applications.

The extensive data access and complex architectures of these models offer unprecedented potential for content creation and transformation. Their ability to learn from diverse sources, generate various multimedia formats, and convert inputs from one format to another opens up a wide array of possibilities in multimedia generation and conversion, making these models indispensable tools in today's technologically advanced world.

In summary, program synthesis is revolutionizing problem-solving by enabling non-experts to

automate solutions without requiring deep knowledge of algorithm design and implementation [14]. Through the strategic use of various constraints, search spaces, and synthesis methods—both deductive and inductive—the field of program synthesis continues to evolve. While the need for complete specifications has historically been a barrier, recent technological advancements now allow for the use of natural language, thereby enhancing the field's accessibility and practical application.

2.2 Natural Language Processing

Artificial intelligence systems have significantly advanced the development of complex cognitive tasks. Natural Language Processing (NLP) serves as a pivotal bridge between human languages and computers, facilitating a myriad of applications [15]. Among the foundational techniques in NLP are regular expressions, which are essential for executing various practical NLP tasks.

Progress in NLP has led to sophisticated approaches for tasks such as text classification, knowledge discovery, and word recommendation. Prominent algorithms for word embedding—such as Word2Vec [16], GloVe [17], and Gensim [18] – play critical roles in these areas by capturing semantic relationships between words and enabling vector representations that are used in downstream tasks. Furthermore, deep learning-based sequence to sequence models (seq2seq) [19] have proven highly effective in machine translation tasks, facilitating the transformation of text from one language to another with high accuracy. Techniques that consider word order and linguistic elements like phonemes and sentences are instrumental in enabling inference and generating novel sentence elements, making models like Transformers especially powerful in generating human-like text [20-21].

Sequence modeling is another critical domain within NLP. Long Short-Term Memory networks (LSTM)[22] are particularly advantageous for these tasks due to their capacity to retain long-term information in a sequence. Unlike recurrent neural networks (RNN), which typically process information through tree structures in a seq2tree fashion, LSTMs incorporate bidirectional flows, thereby enhancing efficiency and performance in a variety of tasks, including speech recognition, time series prediction, and text generation[23]. More recent advancements, such as Bidirectional Encoder Representations from Transformers (BERT), further extend the capabilities of sequence models by pretraining on large corpora and fine-tuning for specific tasks, achieving state-of-the-art results in many NLP benchmarks. In conclusion, NLP technologies enable the seamless integration of natural language understanding within systems, thereby meeting diverse end-user needs and expanding the scope of possible applications. These advancements have profound implications not only in traditional applications like translation and sentiment analysis but also in emerging areas such as conversational AI, content generation, and human-computer interaction, where the ability to understand and generate natural language is crucial. In the next section we will analyze the method we used for this research.

3. Method

To explore diverse perspectives and support the research presented in this work, we adopted the systematic literature review methodology proposed by [24], while also incorporating recommendations from [25–28]. This methodology provides a rigorous framework for exploring the synergies between program synthesis and natural language processing, thereby enriching the research landscape and informing future studies in these areas. Additionally, this research considers fundamental aspects such as requirements, models, input-output formats, and evaluation metrics.

The systematic mapping we employ follows a structured approach that includes Research Questions, Search String, Data Sources, Selection Criteria, and Quality Assessment.

3.1 Research Questions

The objective of our systematic literature review is to obtain a comprehensive understanding of the

key components involved in program synthesis and code generation, particularly through the lens of Natural Language Processing. We aim to reveal the mechanisms underlying automatic program generation and identify areas needing further research to thoroughly understand the context and advantages of program generation via synthesis.

The research questions formulated for this study are designed to systematically dissect these aspects:

Q1. What are the application areas?

Q2. What are the inputs used to synthesize a program?

Q3. What are the outputs generated from the program synthesizer and how are they used?

Q4. What type of synthesis is used?

3.2 Search String

We defined a search string aimed at capturing the intersection of key research domains: [((“Program” OR “Code”) AND (“Synthesis” OR “Generation”)) AND (“Natural Language Processing” OR “NLP”). This string was used to ensure that all pertinent literature was considered. Depending on the database, a general search string was defined and adapted to each search engine.

3.3 Data Sources

Five major data sources were selected to conduct a comprehensive search for literature related to program synthesis and natural language processing. The sources include ACM Digital Library, IEEE Xplore, Springer, Elsevier, and MDPI. These platforms were chosen for their extensive repositories of scientific papers and their relevance to the fields under study.

3.4 Selection Criteria

To ensure a focused and relevant data collection process, several inclusion and exclusion criteria were meticulously applied:

The exclusion criteria eliminated other types of documents, such as unpublished works, books, courses, newspapers, and master’s and doctoral theses.

The inclusion criteria considered only journal articles and conference proceedings published between 2014 and 2024, which allowed us to capture the most recent advances in the field. The search parameters were carefully established to filter data by titles, abstracts, and keywords of journal articles and conference proceedings that met the inclusion criteria. This methodological rigor ensured the collection of the most relevant and beneficial data for our systematic review.

3.5 Quality Assessment

Each study was evaluated using the criteria from the Center for Reviews and Dissemination (CRD) of the University of York, as well as the Database of Abstracts of Reviews of Effects (DARE) [29]. The criteria are based on three quality assessment (QA) questions:

QA1. Are sufficient details about the individual included studies presented?

QA2. Does it provide evidence to answer the research questions for this systematic review?

QA3. Is it a referenced study?

The questions were scored as follows:

- **QA1:** Y (yes), presents sufficient details in the study, P (Partially), presents information partially; N (no), does not have details and cannot be easily inferred.
- **QA2:** Y (yes), The authors based their research in such a way that they included appropriate strategies, identified and made reference to all the journals that addressed the topic of interest, and the study answered all the research questions; P (Partially), The study only partially answered the research questions, N (no), They did not answer the research

questions, and lacked adequate context.

- **QA3:** Y (yes), They are highly referenced studies, P (Partially), The study has a certain number of citations, N (no), The study does not have citations.

The scoring procedure was Y = 1, P = 0.5, N = 0 where information is not specified. Consequently, the possible score that could be obtained for assessing the quality of a primary study was in the range of 0 to 3 points. In this sense, the articles considered had to achieve a rating of 1.5 at least. In the next chapter we will discuss the results obtained.

4. Results

We thoroughly analyzed the full texts of 20 articles that met our selection criteria. Below, we present the initial results obtained from these articles, providing a summary of the studies included in our systematic review. This summary aims to establish a foundational understanding of the scope and impact of the research conducted in the field.

Following the initial overview, we present the quality evaluation of each article to ensure the reliability and validity of the reported findings. This evaluation was essential to maintaining the integrity of our systematic review.

Next, we will demonstrate how we answered the specific research questions posed at the beginning of our study. This analysis will help us identify key trends, gaps in current research, and potential areas for future research, aligning our findings with the overall objectives of our research.

4.1 Summary of the Studies

In this section we present a summary of the works examined, with the goal of answering the research questions formulated. The systematic search across the specified data sources initially yielded a total of 680,924 articles. By applying the inclusion criterion of publication years from 2014 to 2024 and the focus on journals and conferences, the results were refined to 401,104 articles. Further application of criteria related to the research topic narrowed this down to 20 articles directly relevant to the research objectives. The methodology applied in this search is summarized in Fig. 2.

Among these 20 relevant articles, 6 (30%) were published in journals, while 14 (70%) appeared in conference proceedings, as depicted in Fig. 3. We observed a clear trend of increasing publications up until 2021 and 2022, followed by a sharp decline in 2023. This trend could potentially reverse in 2024, influenced by emerging developments in the field.

The research questions and corresponding answers presented in this study have significant implications only 20 papers were directly relevant to the research questions under consideration, as detailed in Table 1.

Among the data sources, the ACM Digital Library demonstrated the highest precision, with an accuracy rate of 0.00215% in yielding relevant articles. A detailed distribution of relevant articles from each source is presented in Table 2. This allows for a discussion on how each identified source contributes to the understanding of program synthesis and natural language processing as outlined by the objective of this paper.

4.2 Quality evaluation

The results of the application of the quality evaluation show that, on average, the studies had a score of 2 points, with the exception of studies S1 and S2 that obtained a score of 1.5. As it can be seen in Table 3, the articles are of good quality and relevant to the investigation.

4.3 Application areas (Q1)

Program synthesis has become a pivotal tool in various domains, demonstrating significant utility, particularly in software engineering. Below we can see the application areas identified in the selected

papers:

- **Program Synthesis for Education:** Enhancing learning and teaching in logic and programming through the generation of code from natural language problems.
- **Program Synthesis for Software Development:** Improving developer efficiency and productivity by automatically generating code, API calls, and optimizing program synthesis tools.
- **Program Synthesis for Robotics and AI:** Simplifying the programming of complex tasks in robotics and generating solutions for various NLP problems through advanced AI models.

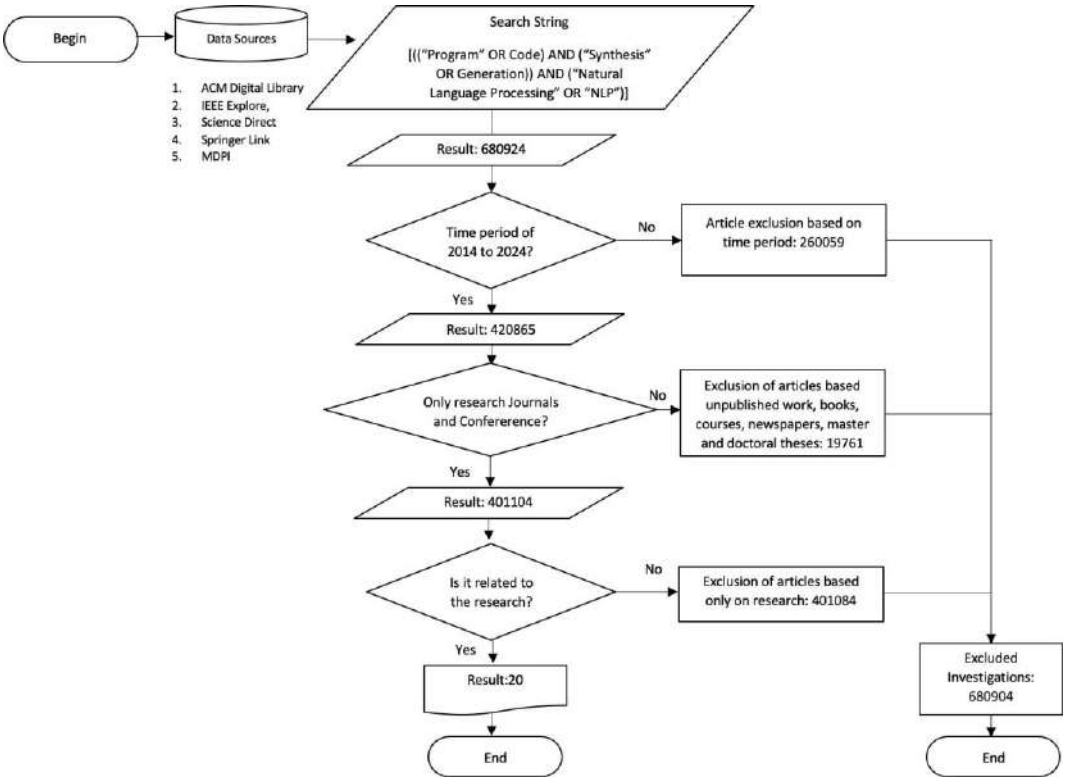


Fig. 2. Steps for the extraction of relevant documents from the selected sources.

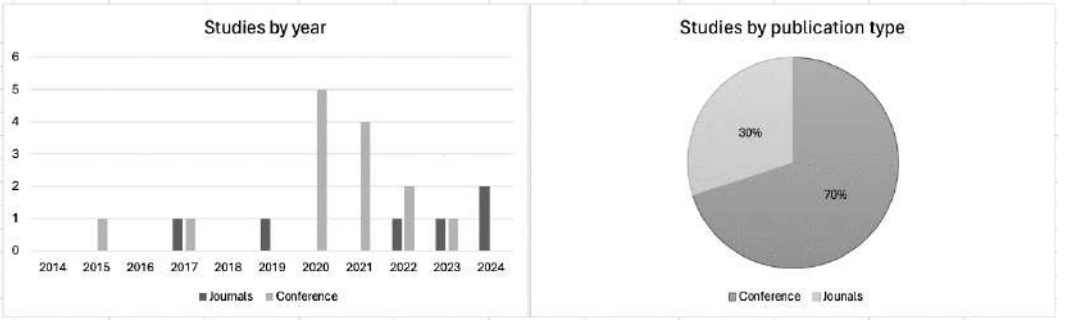


Fig. 3. Type of articles according to inclusion criteria and average of article according to year of publication.

Table 1. Selected Papers.

Title	Author(s)	Year	Ref
S1 Domain specific program synthesis	Archana, P., Harish, P. B., Rajan, N., P, S., and Kumar, N. S.	2021	[30]
S2 Collective intelligence for smarter neural program synthesis	Daiyan. W, Wei. D, and Yating. Z.	2020	[31]
S3 Generating context-aware API calls from natural language description using neural embeddings and machine translation	Phan, H., Sharma, A., and Jannesari, A.	2021	[32]
S4 HISyn: Human Learning-Inspired Natural Language Programming	Nan, Z., Guan, H., and Shen, X.	2020	[33]
S5 Interactive Program Synthesis by Augmented Examples	Zhang, T., Lowmanstone, L., Wang, X., and Glassman, E. L.	2020	[34]
S6 Deep nlp-based co-evolvement for synthesizing code analysis from natural language	Nan, Z., Guan, H., Shen, X., and Liao, C.	2021	[35]
S7 Interactive synthesis of temporal specifications from examples and natural language	Gavran, I., Darulova, E., and Majumdar, R.	2020	[36]
S8 Programming bots by synthesizing natural language expressions into API invocations	Zamanirad, S., Benatallah, B., Barukh, M. C., Casati, F., and Rodriguez, C.	2017	[37]
S9 Egeria – A Framework for Automatic Synthesis of HPC Advising Tools through Multi-Layered Natural Language Processing	Hui. G, Xipeng. S, and Hamid. K.	2017	[38]
S10 Interactive Synthesis using Free-Form Queries	Tihomir. G and Viktor. K.	2015	[39]
S11 Jigsaw – Large Language Models meet Program Synthesis	Naman. J, Skanda. V, Arun. I, Nagarajan. N,	2022	[40]
S12 Many-objective Grammar-guided Genetic Programming with Code Similarity Measurement for Program Synthesis	Ning. T, Anthony. V, and Takfarinas. S.	2023	[41]
S13 Program Synthesis Through Learning the Input-Output Behavior of Commands	Sihyung. L, Seung. Y. Nam, and Jiyeon. K.	2022	[42]
S14 Assessing Similarity-Based Grammar-Guided Genetic Programming Approaches for Program Synthesis	Ning. T, Anthony. V, Takfarinas. K.	2022	[43]
S15 Generative Model for NLP Applications based on Component Extraction	Bhardwaj, P. Khanna, S. Kumar, and Pragya.	2020	[44]
S16 Multi-modal program inference: a marriage of pre-trained language models and component-based synthesis	Kia. R, Mohammad. R, Summit. G and Vu. L.	2021	[45]
S17 Prompt Problems: A New Programming Exercise for the Generative AI Era	Amarouche, B. A. Becker, and B. N. Reeves.	2024	[46]
S18 Automatic Acquisition of Annotated Training Corpora for Test-Code Generation	Magdalena. K and John. D. K.	2019	[47]
S19 Natural Language Generation and Understanding of Big Code for AI-Assisted Programming	Man-Fai, W. Shangxin. G, and Ching-Nam. H.	2023	[48]
S20 Effectiveness of ChatGPT in Coding: A Comparative Analysis of Popular Large Language Models	Carlos. E. C, Mohammed. N. A, and R. K.	2024	[49]

Table 2. Total items extracted.

Data Sources	Result	Useful Articles	Accuracy
IEEE	384457	9	0.00002%
ACM	2781	6	0.00215%
SPRINGER	494	1	0.00202%
ELSEVIER	2708	1	0.00036%
MDPI	10664	3	0.00028%

Table 3. Evaluation of the quality of the studies.

Study	QA1	QA2	QA3	Total Score
S1	P	Y	N	1.5
S2	P	Y	N	1.5
S3	Y	Y	P	2.5
S4	Y	Y	P	2.5
S5	Y	Y	Y	3
S6	Y	Y	P	2.5
S7	Y	Y	P	2.5
S8	P	Y	P	2
S9	Y	Y	N	2
S10	P	P	Y	2
S11	Y	Y	Y	3
S12	P	Y	P	2
S13	Y	Y	N	2
S14	Y	Y	P	2.5
S15	Y	Y	Y	3
S16	Y	Y	Y	3
S17	Y	Y	N	2
S18	Y	Y	P	2.5
S19	Y	Y	Y	3
S20	Y	Y	N	2

Education:

- [30] focuses on the use of program synthesis to solve propositional logic problems in an educational context, emphasizing the generation of code from problems described in natural language. This approach is ideal for teaching and learning in fields related to logic and programming.
- [46] introduces “Prompt Problems” to teach students how to write effective prompts for generating code using large language models (LLMs), helping them develop skills in formulating natural language prompts that produce functional code.

Software Development:

- [31] integrates collective intelligence and bio-inspired algorithms to optimize accuracy in code generation from user intents.
- [32] improves developer efficiency by automatically generating API calls based on natural language descriptions and the context of the surrounding code.
- [33] enhances code generation through natural language understanding, specifically aimed at software development.
- [34] develops interactive program synthesis tools, particularly for creating regular expressions, using augmented examples to clarify user intent and facilitate automatic code generation.
- [35] evaluates the effectiveness of ChatGPT and other large language models in code

generation tasks, highlighting their utility as programming assistance tools.

- [37] focuses on the development of interactive program synthesis tools for creating regular expressions using augmented examples.
- [38] creates a platform called BotBase that allows the transformation of natural language expressions into API invocations, facilitating bot programming.
- [39] creates advisory tools for optimizing high-performance computing programs using natural language processing.
- [40] develops a support tool for IDEs that generates Java code snippets based on free-text queries combining English and code.
- [41] uses pre-trained language models like GPT-3 and Codex to generate code from natural language descriptions, optimized for complex APIs like Python Pandas.
- [42] employs grammar-guided genetic programming for program synthesis, using multiple code similarity measures to improve accuracy in generating code from textual descriptions and input/output examples.
- [43] evaluates and improves the use of grammar-guided genetic programming for program synthesis, guiding the evolutionary process with code similarity measures.
- [45] combines pretrained language models with component-based synthesis techniques to generate programs from natural language descriptions and specific examples, particularly for generating regular expressions and CSS selectors.
- [47] focuses on the automatic creation of annotated data sets to generate automated test cases from quasi-natural language descriptions, using machine learning and machine translation techniques.
- [48] reviews the use of large language models trained with Big Code for various AI-assisted programming tasks, including code generation, completion, translation, refinement, summarization, defect detection, and clone detection.

Robotics:

- [36] facilitates task specification for robots using linear temporal logic (LTL) from natural language examples and interactions, simplifying the programming of complex and specific tasks in robotics applications.

Artificial Intelligence:

- [44] creates a generative model for natural language processing (NLP) applications, extracting meaningful components from case studies to address problems such as reading text, interpreting speech, measuring sentiment, and determining important parts, generating optimized solutions for different NLP problems.

4.4 Inputs used to synthetize a program (Q2)

In this section, the primary focus is to identify the different types of inputs that will be processed by the synthesizer programs. The exploration of the literature has allowed us to identify how these works take natural language expressions and synthesize examples based on the user's intended purpose, using different techniques to achieve the various objectives proposed by the authors. The results of relevant articles are detailed in Table 4.

The study by [30] introduces a tool for end-user programming designed to simplify the programming process and enable programmers to focus more on the core logic of the program. This tool removes the need to deal with language syntax and other domain-specific aspects. User input is provided in the form of a propositional verbal problem, which consists of facts, conditionals, and questions, thereby establishing the basis for a learning approach.

[31] centers on the automatic generation of source code from various user intents. The authors utilized natural language task descriptions as inputs, enabling the identification of web page tags that align with these characteristics. This study demonstrates the versatility of user intent expression and represents significant progress in solving programming tasks based on natural language descriptions, requiring minimal information about the target program.

Table 4: Types of inputs from different examples of program synthesis using NL.

Types of inputs	Articles
Verbal problems (Query)	[30]
Natural language task descriptions	[31], [48], [46]
Sentences and a part of the surrounding context.	[32], [39]
Natural language (query)	[33], [38], [45]
Description of a method in NL	[34], [47]
NL queries based on dependency structure	[35]
Specific descriptions	[36], [40]
Short description of a specification	[37], [42]
High-level specifications	[41]
Textual Problem Descriptions	[43], [44]
Programming Prompts	[49]

Similarly, the study in [32] employed a unique method involving the pairing of an instruction sentence with a section of corresponding code. The input consisted of a natural language user intent and a drafted method, using the Java language. A method name generator was then employed to extract tokens and variable names from natural language descriptions and adjacent code tokens, thus predicting potential method names.

The research presented in [33] adopts an approach driven by natural language understanding. The input consists of a natural language query containing a list of synonyms, named entities, and a dictionary of prepositions. This method reduces the need for extensive labeled examples, thereby freeing users from the task of gathering examples and facilitating natural language programming, especially in domains where labeled examples are difficult to obtain. The study in [34] investigates the use of natural language descriptions of methods as input to improve concrete word recognition. The researchers introduce a semantic analyzer that links variables to specific operational information, thus describing the method's particular behavior, parameter name, and return value information.

The field of code analysis presents numerous complexities, especially those associated with data types and operations. The research in [35] introduces a tool that significantly mitigates these complexities. This tool leverages natural language queries, drawing upon dependency structures in language, to interpret the code. The tool specifically automates the analysis of asymmetric binary relations between words in a sentence, such as subordinate words and their dependencies. In other words, it uses the syntactic structures of natural language to build a semantic understanding of code. This approach not only aids in extracting the core meaning of the code but also makes the process more comprehensible and accessible to programmers.

Simultaneously, natural language descriptions and programming by example have emerged as "user-friendly" alternatives for specifying complex tasks. [36] addresses these issues by using specific descriptions as inputs. This method generates grammatical rules for producing parseable commands, thus facilitating the straightforward specification of complex, repetitive tasks.

Lastly, although modern bot creation systems detect user intent, they require considerable development and configuration effort for each use case. [37] introduces a tool that uses a concise

specification description as input, assisting in the generalization of critical tasks in the program generation process.

The inputs of Egeria [38] include optimization guides or other domain-specific documents relevant to HPC. Additionally, user queries or performance profiling reports can be fed into the synthesized advising tool to receive specific optimization advice.

The inputs used by the synthesis tool proposed in [39] include free-form queries composed of a mixture of English and Java code. These queries can describe desired functionalities or operations in natural language, possibly combined with partial code snippets. The system also incorporates context from the developer's current work in the IDE, such as the cursor position and existing code, to better understand and generate the appropriate code fragments.

Jigsaw [40] accepts multi-modal inputs for synthesizing programs. Users can input their intent or requirements in natural language and also include test cases, input/output examples. These are used to further specify the intended functionality of the code, helping to refine the synthesis process and ensure that the generated code meets the user's needs.

The inputs for synthesizing a program using MaOG3P [41] include high-level specifications or textual descriptions of the desired functionality of the program. Particularly, input/output examples specify what the program should produce given certain inputs, helping to guide the genetic programming process to evolve correct and efficient code.

The inputs for the program synthesis system proposed by [42] take the form of short descriptions of specifications. The system understands the available commands and their syntax, which guide the synthesis process. For instance, the system uses examples of desired inputs to learn and generate the corresponding program.

The inputs used in [43] include textual problem descriptions that describe a programming task provided in natural language, grammatical specifications, such as a defined grammar that dictates the syntax of the programming language in which the programs are developed, and similar code that is used to evaluate the suitability of evolved programs against a target source code, improving the relevance of the generated programs for the given problem descriptions.

The inputs for the NLP generative model discussed in [44] take the form of a Textual Problem Description, which is a description provided in natural language that outlines the problem to be solved by the model. These descriptions are extracted from case studies that identify significant components relevant to the problem being addressed.

The inputs for synthesizing programs in [45] take the form of natural language queries. For example, students craft prompts in natural language that describe the desired functionality or outcome of a program. This kind of input helps define the problem that needs to be solved by the generated code, guiding the LLM towards appropriate solutions.

The input for the multi-modal synthesis approach described in [46] is a Natural Language Description, which is a broad, often ambiguous description of a desired functionality. This kind of input provides a specification of how the desired code should function.

The inputs of the synthesis process described in [47] are descriptive method names, which are extracted from source code and are used as natural language descriptions of the functionality of the code, and also function bodies that are aligned with the method names to form a parallel text code corpus.

The inputs used in [48] are natural language descriptions, which describe the desired functionality in natural language, and also existing code fragments that serve as context or examples for the desired operations.

In [49] the inputs used to synthesize programs are programming prompts, which describe what the generated code should accomplish, as well as code examples, that can be used to guide the AI in generating appropriate code structures.

4.5 Outputs generated from the program synthesis (Q3)

Program synthesis offers flexibility by utilizing incomplete specifications, regardless of the specific approach employed, to generate code. The objective is to achieve a degree of final completeness in the produced output. However, it is important to note that the generated output may not always align with the end user's expectations.

[30] leverages postfix expressions (Boolean Logic) to establish a foundation for a domain-independent learning approach to problem-solving via program synthesis concepts. This process enables users, particularly programmers, to streamline their efforts by focusing on the core logic of the program, thereby mitigating concerns about language syntax and other domain-specific elements. Given the input "*Did Mary and Ram go to school?*", the output is "*Cannot be determined / True*".

The development of large and complex software projects requires a workforce trained in the fundamental structures of the programming languages they use. One potential approach to automate this process is the generation of a common keyword list. In this scenario, programmers need not memorize the keyword vocabulary or understand their exact implementation to write a program in the given language. For instance, a list of expected method names could be derived from a method description with surrounding code [32]. For example, for the input "*return random number with max value iterationWeight for Random*", the output would be new "*Random().nextInt(iterationWeight)*".

Alternatively, understanding how programmers code is a complex process that demands practical solutions. By deeply processing programmers' intentions and API documents written in natural language, it is possible to leverage a profound understanding through program synthesis tailored for this specific purpose. This approach circumvents the need for a large number of labeled examples, thus alleviating the user's task of collecting or generating examples. It also significantly impacts traditional methods. For instance, from the input "*Find statements whose init portion declares a single variable which is initialized to the integer literal 0*", the following code (in a DSL) is generated:

```
forStmt(
    hasLoopInit(
        declStmt(
            hasSingleDecl(
                varDecl(
                    hasInitializer(
                        integerLiteral(
                            equals(0)))))))
```

Code library functions have significantly increased developers' programming efficiency. They do so by simplifying constraint generation and accelerating constraint resolution through the creation of complete code based on constraint models of Java classes [34]. A pertinent example is a code fragment in a tree structure, as shown below:

```
(define-fun result () Int (- 1))
(define-fun this ()
    (Seq String) (seq.unit ""))
(define-fun or () String "")
```

This example includes encapsulated functions that streamline and speed up constraint generation through the use of generated constraint models.

Concurrently, attaining high software quality controls is a complex task. It requires support from various program optimizations, software debugging, security measures, and more. Therefore, code analysis in the early stages of development can provide developers with various preemptive options [35]. Such an approach employs "final comparison expressions" that originate from specific natural language descriptions and assist general programmers in conducting automated program analysis.

For instance, given the input "*Find all C++ call expression of the C++ method named string1*", the generated output expression in the form of an AST is:

```
cxxMemberCallExpr(
    callee(
        cxxMethodDecl(
            hasName(string1))))
```

The correct use of specifications often poses a challenge to non-expert users. Therefore, providing an output that illustrates a synthesized specification derived from an example and a natural language description can significantly enhance the accuracy of the synthesis method. Furthermore, it paves the way for the generalization of synthesized tasks to other unseen tasks [36]. For instance, for the expression "*step into water and then visit (6, 4)*", it is possible to obtain an LTL specification as "*step into water and then visit (Num, Num)*".

Undeniably, there are numerous endeavors aimed at refining the process of automatic code generation. Each study provides a perspective on how productivity in development can be enhanced. One increasingly popular approach is the use synthesize API calls from expressions in NL. To fully harness the potential of this approach, [37] propose a tool designed to foster the development of intuitive software solutions. This tool bridges the gap between user needs, expressed in natural language, and API invocations capable of satisfying these needs. An example is: synthesize API calls from expressions in NL

```
<url:https://api.yelp.com/v2,
    path:/search parameters:
    term=[italian,cafes],
    location=[sydney.opera_house]>
```

The outputs generated by Egeria [38] include an advising tool that provides a list of essential rules extracted from the input documents. This tool also serves as a question-answer agent that offers specific optimization suggestions based on user queries or performance profiling reports. Fig. 4 shows an example rule that is used to guide programmers in optimizing code more effectively without needing to manually sift through extensive documentation.

```
if(tx % 2 == 0 && ty % 2 == 0)
    out[tx * width + ty] = 2.0 * in[tx * width + ty]/sum;
else if(tx % 2 == 1 && ty % 2 == 0)
    out[tx * width + ty] = in[tx * width + ty]/sum;
else if(tx % 2 == 1 && ty % 2 == 1)
    out[tx * width + ty] = (-1.0) * in[tx * width + ty]/sum;
else
    out[tx * width + ty] = 0.0f;
```

Fig. 4. The Optimized Block [38].

The outputs generated by [39] are Java code fragments that respect Java syntax, type, and scoping rules, as well as conform to common usage patterns derived from a statistical analysis of existing code. These code fragments are presented to the developer within the IDE, offering several ranked suggestions that the developer can choose from. The primary use of these outputs is to insert appropriate code snippets into the developer's project, helping to bridge the gap between a high-level concept expressed in natural language and executable Java code.

The output of Jigsaw [40] is executable code that matches the user's specified intent and passes given test cases. Fig. 5 shows code that is generated after processing through a series of program analysis and synthesis techniques, which include correcting common errors detected in the initial outputs from pre-trained language models (PTLMs) like GPT-3 or Codex. The generated code helps programmers quickly implement solutions and focus on higher-level design and problem-solving tasks rather than the nuances of specific API calls or syntax correctness.

The outputs generated by MaOG3P [41] are executable code snippets that meet the requirements

specified through the input descriptions and examples. These outputs are used to automate coding tasks, reduce development time, and improve the efficiency of the programming process. By synthesizing code that satisfies both the syntactic and semantic correctness, the generated programs help developers by providing ready-to-use code snippets that can be integrated into larger projects or used as standalone solutions.

Code Before	Code After
`out=data[data.index.isin(test.index)]`	`out=data[~data.index.isin(test.index)]`
`df=df[df['foo']>70)|df['foo']<34]`	`df=df[(df['foo']>70)|(df['foo']<34)]`
`out=df.iloc[0,"HP"]`	`out=df.loc[0,"HP"]`
`dfout=df1.append(df2,ignore_index=True)`	`dfout=df1.append(df2)`
`dfout=dfin.duplicated()`	`dfout=dfin.duplicated().sum()`
`train=data.drop(test)`	`train=data.drop(test.index)`
`dfin=dfin["A"].rolling(window=3).mean()`	`dfin["A"]=dfin["A"].rolling(3).mean()`
`dfout=dfin[(x<40)|(y>53)&(z==4)]`	`dfout=dfin[(x<40)|(y>53)&(z==4)]`

Fig. 5. Applications (Code After) of learned transformations on code snippets produced by PTLM (Code Before) [40].

The output from this system [42] is an executable program that conforms to the specifications derived from the input-output examples provided. These programs can then be used directly within software applications, helping to automate tasks or improve software functionality with minimal human coding effort.

The outputs from the program synthesis approach proposed in [43] are executable pieces of code that align with user-defined specifications and grammar rules. Fig. 6 shows programs that are evaluated for similarity against target codes to ensure that they meet the specified requirements. This can be used in Software Development to automate or speed up the development process by providing ready-to-use code snippets that fit the user's intent. Finally, this is an example of teaching tools to demonstrate various programming techniques and solutions.

Problem	Textual Description	# Input/Output Pair	
		Training	Testing
Number IO	Given an integer and a float, print their sum.	25	1000
Median	Given 3 integers, print their median.	100	1000
Smallest	Given 4 integers, print the smallest of them.	100	1000

Fig. 6. Representation of target programs [43].

The outputs of the NLP generative model [44] are optimized solutions for NLP tasks. The model generates solutions that address specific NLP-related problems like speech interpretation, sentiment analysis, and text processing and adapted responses, because the system uses the outputs to adapt its responses based on the input it receives, making it suitable for interactive applications such as virtual assistants.

In the case of [45], the outputs are generated code based on prompts provided. A LLM is used to generate code that attempts to solve a specified problem, and then the generated code is evaluated against test cases to determine its correctness. This process aids in learning by providing immediate feedback on the effectiveness of the prompt and the functionality of the code.

The outputs generated in [46] are executable code snippets that precisely match the combined specifications provided by the natural language descriptions and the examples. Fig. 7 shows how that works. The synthesized programs are used in software development to automate coding tasks, ensuring that the generated code meets both broad functional requirements and specific operational details.

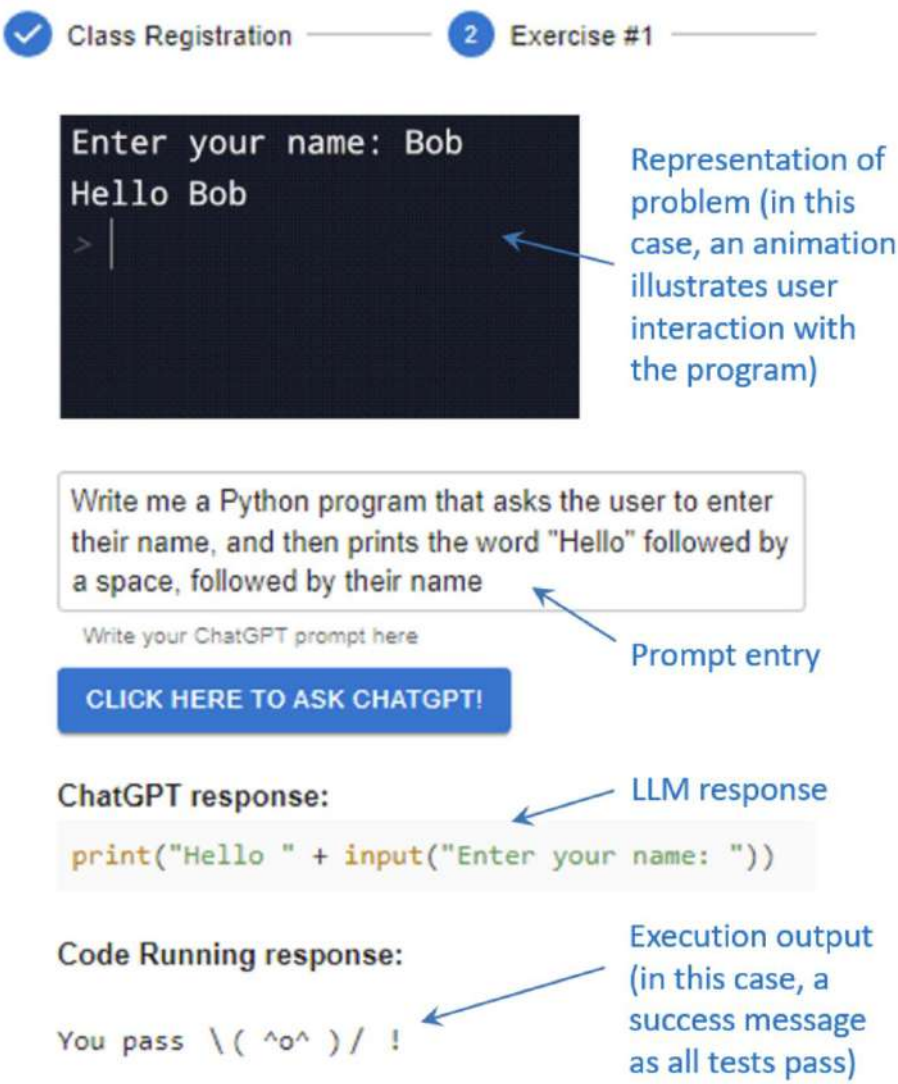


Fig. 7. Interface layout for a Prompt Problem within the web-based Promptly tool [46].

The outputs generated [47] are code fragments. They consist of test function names (as natural language descriptions) aligned with their respective function bodies (as code). These examples are compiled and semantically relevant test cases generated by machine learning models trained on the synthesized corpora. The goal is automating the creation of unit tests, reducing the time and effort required for manual test case development.

The outputs generated by the program synthesizer in [48] include executable code, which can be run directly or integrated into larger software projects. The study also allows generating descriptions in natural language or summaries of code blocks, useful for documentation. These results are used to improve productivity, improve code quality, help maintain and document code, and ensure the reliability and security of software systems.

The outputs from the program synthesizer presented in [49] are executable code within a software environment, as seen in Fig. 8. These results are used to directly implement functional requirements in software projects, automate routine coding tasks, and improve overall software quality and reliability through improved error detection and resolution capabilities.

4.6 Type of program synthesis used (Q4)

Program synthesis employs distinct methods, primarily classified into deductive and inductive synthesis, each tailored to specific aspects of programming from formal specifications.

Deductive Program Synthesis involves formally deriving a program from a given specification, treating the process as a theorem-proving task. It necessitates demonstrating that a program's output satisfies the conditions set out in the specification, a process confirmed through theorem proof. This approach requires detailed, often complex specifications and is particularly useful when correctness is paramount. In practice, deductive synthesis involves the use of formal methods to ensure that every step in the program construction adheres strictly to the logical constraints outlined in the specification. Notable works in this domain include [35], [34], and [37], which explore various applications of theorem proving in program synthesis, demonstrating the effectiveness of this approach in generating highly reliable software.

```
Task_id = "162"
prompt = """Write a function to calculate the sum (n - 2*i) from i=0 to n // 2,
        for instance n + (n-2) + (n-4)... (until n-x =< 0)."""
test_list = [
    "assert sum_series(6) == 12",
    "assert sum_series(10) == 30",
    "assert sum_series(9) == 25",
]
def sum_series(n):
  """Calculates the sum of the series (n - 2*i) from i=0 to n // 2.
  Args:
    n: The upper limit of the series.
  Returns:
    The sum of the series.
  """
  sum = 0
  for i in range(n // 2):
    sum += n - 2 * i
  return sum
```

Fig. 8. Example of outputs obtained on a data set of basic Python problems (MBPP) [49].

Inductive Program Synthesis, in contrast, starts from incomplete problem descriptions, which might include test cases, characteristics of desired and undesirable software behaviors, input-output examples, or computational traces. This approach encompasses several methodologies that aim to generalize from these examples to produce a program that satisfies the specification in a broader sense:

- *Genetic Programming and Incremental Evolution*, as exemplified by [30] where the synthesis process evolves programs iteratively, optimizing them to better fit the examples provided.
- *Counterexample-Guided Inductive Synthesis*, explored in studies such as [32] and [36], refines candidate programs by iteratively correcting them based on counterexamples, thus gradually improving their correctness.
- *Neural Program Synthesis*, with key contributions from [33] and [31], leverages deep learning models to synthesize programs from natural language or other high-level inputs, demonstrating significant advancements in automating complex programming tasks.

Egeria [38] utilizes an unsupervised, multi-layered design leveraging NLP techniques. Although not explicitly categorized, its synthesis approach suggests inductive reasoning through optimization based on general guidelines and specific user queries.

Jigsaw [40] integrates inductive synthesis with corrective transformations, initially using pre-trained language models for generating code snippets from natural language inputs, followed by corrective transformations to ensure accuracy, blending inductive learning with deductive refinements.

MaOG3P [41] and the approach outlined in [42] emphasize inductive synthesis through genetic programming and machine learning, respectively, focusing on evolving programs to meet specific input-output behaviors based on learned patterns.

The synthesis methodologies in [44] and [45] also follow inductive approaches, generalizing from specific examples to create applicable solutions across new scenarios.

Lastly, the approaches in [47], [48], and [49] exemplify the inductive synthesis prevalent in AI-assisted programming, where large datasets of code are used to predict and generate new code segments, demonstrating how modern AI tools, like ChatGPT, generalize from extensive training data to produce functional programming solutions.

This study concludes with an examination of [46], which combines inductive and deductive elements. The process starts with PTMs generating initial code candidates, followed by a Component-Based Synthesis (CBS) approach that deductively constructs the final program, ensuring it meets the provided examples through systematic component assembly and refinement. In the next chapter we present our main discussions of the study.

5. Results discussion

In this section the results of this systematic literature review reveal both the progress and ongoing challenges in the field of program synthesis, particularly when interfacing with natural language processing (NLP). The analysis of 20 selected studies highlights several key trends and areas of focus that have emerged over the past decade, also the systematic literature review on program synthesis and natural language processing (NLP) reveals significant advancements and emerging trends in this field. A key finding is the increasing integration of advanced artificial intelligence models, especially large language models (LLMs), which have demonstrated remarkable capabilities in interpreting natural language specifications and generating executable code.

This development is democratizing software development, allowing users with little or no programming experience to create functional applications using natural language instructions. The review also highlights the evolution of program synthesis methodologies, which have transitioned from rule-based approaches to more sophisticated techniques that leverage machine learning and genetic programming. These modern techniques can learn from input-output examples and user interactions, thus improving the accuracy and efficiency of code generation. However, significant challenges remain, such as achieving high precision in interpreting complex natural language requirements and ensuring responsible AI practices to guarantee the reliability of the generated code.

On the other hand, ambiguity in natural language specifications and the scalability of program synthesis systems represent crucial challenges in automatic code generation. Ambiguity, inherent in natural language, can lead to multiple interpretations of the same instruction, making it difficult to correctly understand and translate the user's intentions into executable code. To mitigate this problem, it is necessary to develop techniques that effectively disambiguate specifications, using contextual models and interactive visualization tools. On the other hand, scalability is essential for these systems to be able to handle complex tasks and large volumes of data without losing performance. This requires the implementation of optimizations such as parallel processing and model compression, ensuring that systems can adapt to various domains and contexts without compromising the quality of the generated code.

Finally, the potential applications of program synthesis go beyond traditional software development. In the educational field, program synthesis tools are used to teach logic and programming concepts,

automatically generating code from problem descriptions provided in natural language, making programming more accessible to a broader audience.

6. Conclusions

This investigation has examined the state of program synthesis from natural language, uncovering various trends and motivations within the field of automatic code generation. Through meticulous analysis of current literature, this study underscores the expanding role of natural language processing (NLP) tools and their potential to profoundly influence computing disciplines.

The advancements in NLP not only enhance communication capabilities but also facilitate the creation of sophisticated methods for generating syntactic representations of programming languages, as highlighted in the referenced paper [50]. Such methodologies leverage pre-trained, language-based components, promising to refine the process of transforming human language into executable code.

Furthermore, with AI-based systems becoming ever more integral to daily life and the disruptive capabilities of generative AI models, the incorporation of responsible AI practices becomes imperative. This approach will ensure the development and deployment of large language models and other generative systems are both reliable and trustworthy, fostering greater confidence in their applications.

Future research in the field of program synthesis should focus on improving the interpretability of systems, allowing coding decisions to be more understandable and reliable, especially for non-expert users. Furthermore, domain-specific synthesis models should be developed, using specialized datasets to improve the accuracy and relevance of synthesized programs. Optimizing the scalability and computational efficiency of these systems is equally vital, ensuring their large-scale adoption. Finally, it is critical to incorporate ethical considerations and responsible artificial intelligence principles, ensuring fairness, accountability, and transparency in synthesis systems, and minimizing biases. As for practical implications, integrating program synthesis tools into educational platforms can facilitate learning programming, while in software development, automating repetitive tasks and codebase generation will allow developers to focus on more creative aspects. Furthermore, improving the accessibility and usability of applications through natural language interfaces driven by program synthesis could revolutionize human-computer interaction.

Overall, this study illuminates the dynamic field of program synthesis from natural language, advocating for continued research and development. By harnessing advanced NLP and responsible AI, the gap between human language and computer programming can be bridged more effectively, setting a foundation for future innovations in automatic code generation.

References

[1]. J. Fu, F. B. Bastani, and I.-L. Yen, in Innovations for Requirement Analysis. From Stakeholders' Needs to Formal Designs, edited by B. Paech and C. Martell (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007), pp. 43–61, ISBN 978-3-540-89778-1.

[2]. Z. Manna and R. J. Waldinger, Toward automatic program synthesis (1971), URL https://doi.org/10.1145/362566. 362568.

[3]. S. Jiho and N. Jaechang, A survey of automatic code generation from natural language (2021).

[4]. A. Bandi, P. V. S. R. Adapa, and Y. E. V. P. K. Kuchi, The power of generative ai: A review of requirements, models, input–output formats, evaluation metrics, and challenges (2023), URL https://www.mdpi.com/1999-5903/15/8/260.

[5]. S. Gulwani, in Formal Methods in Computer Aided Design (2010), pp. 1–1.

[6]. Z. Manna and R. Waldinger, A deductive approach to program synthesis (1980), URL https://doi.org/10.1145/357084. 357090.

[7]. E. Kitzelmann, in Approaches and Applications of Inductive Programming, edited by U. Schmid, E. Kitzelmann, and R. Plasmeijer (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010), pp. 50–73, ISBN 978-3-642-11931-6.

[8]. A. F. Subahi, Cognification of program synthesis—a systematic feature-oriented analysis and future direction (2020), URL https://www.mdpi.com/2073-431X/9/2/27.
[9]. J. Sauvola, S. Tarkoma, M. Klemettinen, J. Riekki, and D. Doermann, Future of software development with generative ai (2024).
[10]. C. Ebert and P. Louridas, Generative ai for software practitioners (2023).
[11]. K. S. Kaswan, J. S. Dhatterwal, K. Malik, and A. Baliyan, in 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI) (2023), pp. 699–704.
[12]. G. Brunner, A. Konrad, Y. Wang, and R. Wattenhofer, Midi-vae: Modeling dynamics and instrumentation of music with applications to style transfer (2018), URL https://api.semanticscholar.org/CorpusID:49317433.
[13]. T. Karras, S. Laine, and T. Aila, A style-based generator architecture for generative adversarial networks (2021), URL https://doi.org/10.1109/TPAMI.2020.2970919.
[14]. D. Jonassen, J. Howland, J. L. Moore, and R. Marra, in Learning to Solve Problems with Technology: A Constructivist Perspective (2002).
[15]. E. D. Liddy, Natural language processing. (2001).
[16]. D. Jatnika, M. A. Bijaksana, and A. A. Suryani, Word2vec model analysis for semantic similarities in english words (2019), the 4th International Conference on Computer Science and Computational Intelligence (ICCSCI 2019) : Enabling Collaboration to Escalate Impact of Research Results for Society, URL https://www.sciencedirect.com/science/article/ pii/S1877050919310713.
[17]. Venkatesh, S. U. Hegde, Z. A. S, and N. Y, in 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) (2021), pp. 1–8.
[18]. L. Wang, Y. Ling, Z. Yuan, M. Shridhar, C. Bao, Y. Qin, B. Wang, H. Xu, and X. Wang, Gensim: Generating robotic simulation tasks via large language models (2024), 2310.01361.
[19]. A. Namboori, S. Mangale, A. Rosenbaum, and S. Soltan, Gemquad : Generating multilingual question answering datasets from large language models using few shot learning (2024), 2404.09163.
[20]. T. Mikolov, M. Karafiat, L. Burget, J. Cernocky´, and S. Khudanpur, in Recurrent neural network based language model (2010), vol. 2, pp. 1045–1048.
[21]. T. Mikolov, A. Deoras, D. Povey, L. Burget, and J. Cernocky, Strategies for training large scale neural network language models (2011).
[22]. S. Islam, H. Elmekki, A. Elsebai, J. Bentahar, N. Drawel, G. Rjoub, and W. Pedrycz, A comprehensive survey on applications of transformers for deep learning tasks (2023), 2306.07303.
[23]. P. Yin and G. Neubig, in A Syntactic Neural Model for General-Purpose Code Generation (2017), pp. 440–450.
[24]. B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, Information and Software Technology 51, 7 (2009), ISSN 0950-5849, special Section – Most Cited Articles in 2002 and Regular Research Papers, URL http: //www.sciencedirect.com/science/article/pii/S0950584908001390.
[25]. K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, in Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (BCS Learning and Development Ltd., 2008), EASE'08, p. 68–77.
[26]. N. S. M. Yusop, J. Grundy, and R. Vasa, IEEE Transactions on Software Engineering 43, 848 (2017).
[27]. R. Moguel-S´anchez, C. Mart´ınez-Palacios, J. Ochar´an-Hern´andez, X. Lim´on, and A. S´anchez-Garc´ıa, Programming and Computer Software 49, 712 (2024).
[28]. P. O. Silva-Vasquez, V. Y. Rosales-Morales, and E. Ben´ıtez-Guerrero, Program. Comput. Softw. 48, 685–701 (2022), ISSN 0361-7688, URL https://doi.org/10.1134/S0361768822080187.
[29]. U. Dissemination, The database of abstracts of reviews of effects (dare) (2002).
[30]. P. Archana, P. B. Harish, N. Rajan, S. P, and N. S. Kumar, in 2021 Asian Conference on Innovation in Technology (ASIANCON) (2021), pp. 1–8.
[31]. D. Wang, W. Dong, and Y. Zhang, in 2020 35th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW) (2020), pp. 98–104.
[32]. H. Phan, A. Sharma, and A. Jannesari, in 2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW) (2021), pp. 219–226.
[33]. Z. Nan, H. Guan, and X. Shen, HISyn: Human Learning-Inspired Natural Language Programming (Association for Computing Machinery, New York, NY, USA, 2020), p. 75–86, ISBN 9781450370431, URL https://doi.org/10.1145/3368089. 3409673.
[34]. Z. Zhang, S. Wu, R. Jiang, M. Pan, and T. Zhang, in Proceedings of the Tenth Asia-Pacific Symposium on Internetware (Association for Computing Machinery, New York, NY, USA, 2018), Internetware '18, ISBN 9781450365901, URL https: //doi.org/10.1145/3275219.3275229.

[35]. Z. Nan, H. Guan, X. Shen, and C. Liao, in Proceedings of the 30th ACM SIGPLAN International Conference on Compiler Construction (Association for Computing Machinery, New York, NY, USA, 2021), CC 2021, p. 141–152, ISBN 9781450383257, URL https://doi.org/10.1145/3446804.3446852.
[36]. I. Gavran, E. Darulova, and R. Majumdar, Interactive synthesis of temporal specifications from examples and natural language (2020), URL https://doi.org/10.1145/3428269.
[37]. S. Zamanirad, B. Benatallah, M. C. Barukh, F. Casati, and C. Rodriguez, in 2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE) (2017), pp. 832–837.
[38]. H. Guan, X. Shen, and H. Krim, in SC17: International Conference for High Performance Computing, Networking, Storage and Analysis (2017), pp. 1–14.
[39]. T. Gvero and V. Kuncak, in 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering (2015), vol. 2, pp. 689–692.
[40]. N. Jain, S. Vaidyanath, A. Iyer, N. Natarajan, S. Parthasarathy, S. Rajamani, and R. Sharma, in 2022 IEEE/ACM 44th International Conference on Software Engineering (ICSE) (2022), pp. 1219–1231.
[41]. N. Tao, A. Ventresque, and T. Saber, in 2023 IEEE Latin American Conference on Computational Intelligence (LA-CCI) (2023), pp. 1–6.
[42]. S. Lee, S. Y. Nam, and J. Kim, Program synthesis through learning the input-output behavior of commands (2022).
[43]. N. Tao, A. Ventresque, and T. Saber, in Optimization and Learning – 5th International Conference, OLA 2022, Syracuse, Sicilia, Italy, July 18-20, 2022, Proceedings, edited by B. Dorronsoro, M. Pavone, A. Nakib, and E.-G. Talbi (Springer, 2022), vol. 1684 of Communications in Computer and Information Science, pp. 240–252, ISBN 978-3-031-22039-5, URL https://doi.org/10.1007/978-3-031-22039-5_19.
[44]. A. Bhardwaj, P. Khanna, S. Kumar, and Pragya, Generative model for nlp applications based on component extraction (2020), international Conference on Computational Intelligence and Data Science, URL https://www.sciencedirect. com/science/article/pii/S1877050920308577.
[45]. K. Rahmani, M. Raza, S. Gulwani, V. Le, D. Morris, A. Radhakrishna, G. Soares, and A. Tiwari, Multi-modal program inference: a marriage of pre-trained language models and component-based synthesis (2021), URL https://doi.org/10. 1145/3485535.
[46]. P. Denny, J. Leinonen, J. Prather, A. Luxton-Reilly, T. Amarouche, B. A. Becker, and B. N. Reeves, in Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (Association for Computing Machinery, New York, NY, USA, 2024), SIGCSE 2024, p. 296–302, ISBN 9798400704239, URL https://doi.org/10.1145/3626252.3630909.
[47]. M. Kacmajor and J. D. Kelleher, Automatic acquisition of annotated training corpora for test-code generation (2019), URL https://www.mdpi.com/2078-2489/10/2/66.
[48]. M.-F. Wong, S. Guo, C.-N. Hang, S.-W. Ho, and C.-W. Tan, Natural language generation and understanding of big code for ai-assisted programming: A review (2023), URL https://www.mdpi.com/1099-4300/25/6/888.
[49]. C. E. A. Coello, M. N. Alimam, and R. Kouatly, Effectiveness of chatgpt in coding: A comparative analysis of popular large language models (2024), URL https://www.mdpi.com/2673-6470/4/1/5.
[50]. M. Rabinovich, M. Stern, and D. Klein, Abstract syntax networks for code generation and semantic parsing (2017), URL https://arxiv.org/abs/1704.07535.

Информация об авторах / Information about authors

Роландо РАМИРЕС-РУЭДА учится в аспирантуре по программированию в Университете Веракруса (Мексика). Лектор отделения системного программирования в одном из институтов этого университета. Сфера научных интересов: автоматический вывод, искусственный интеллект, мультиагентные системы.

Rolando RAMÍREZ-RUEDA – PhD Student in Computer Science from the University of Veracruz in Mexico. Professor in the systems department of Veracruz University Institute in Mexico. Research interests: Automated Reasoning, Artificial Intelligence, Multiagent Systems.

Эдгард БЕНИТЕС-ГУЭРРЕРО имеет степень PhD по программированию Гренобльского университета (Франция). Профессор факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных интересов: человеко-машинное взаимодействие,

искусственный интеллект, совместные вычисления, управление данными и визуализация данных.

Edgard BENÍTEZ-GUERRERO – PhD in Computer Science from the University of Grenoble in France. Professor at the Faculty of Statistics and Informatics of the University of Veracruz in Mexico. Research interests: Human Computer Interaction, Artificial Intelligence, Collaborative Computing, Data Management and Visualization.

Кармен МЕЗУРА-ГОДОЙ получил степень PhD по программированию в Университете Савойи (Франция). Профессор факультета статистики и информатики Университета Веракруса (Мексика). Основные научные интересы: человеко-машинное взаимодействие, пользовательский опыт взаимодействия, совместная работа с поддержкой компьютера, визуализация и мультиагентные системы.

Carmen MEZURA-GODOY – PhD in Computer Science from the University of Savoie in France. Professor at the Faculty of Statistics and Informatics of the University of Veracruz in Mexico. Main research interests: Human-Computer Interaction, User eXperience-UX, Computer-Supported Collaborative Work, Visualization and Multiagent Systems

Эверардо БАРСЕНАС имеет степень PhD по программированию Гренобльского университета, работает в должности доцента на факультете программной инженерии Национального университета Мексики. Его научные интересы включают модальную логику, теорию доказательств, автоматизированный логический вывод, описательную логику, проверки моделей и формальные верификации.

Everardo BARCENAS holds a PhD in Computing Science from the University of Grenoble, and is an Assistant Professor in the Computing Engineering Department of the National University of Mexico. His research interests include: Modal Logics, Proof Theory, Automated Reasoning, Description Logics, Model Checking and Formal Verification.

DOI: 10.15514/ISPRAS-2024-36(6)-5

Project Management Maturity Models: A Systematic Review

[1] *J. F. Ruiz-Lopez, ORCID: 0009-0003-9888-8537 <d19ce063@cenidet.tecnm.mx>*
[1] *J. Ortiz-Hernandez, ORCID: 0000-0002-6481-1537 <javier.oh@cenidet.tecnm.mx>*
[2] *E. Bonjour, ORCID: 0000-0001-7851-1095 <eric.bonjour@univ-lorraine.fr>*
[3] *J.-P. Micaelli, ORCID: 0000-0002-9403-1419 <jean-pierre.micaelli@univ-lyon3.fr>*
[1] *Y. Hernandez, ORCID: 0000-0002-8842-0899 <yasmin.hp@cenidet.tecnm.mx>*

[1] *National Technology of Mexico, National Center of Research and Technological Development, Cuernavaca, México.*
[2] *Research Team on Innovative Processes, University of Lorraine, Nancy, France.*
[3] *Magellan Research Laboratory, Lyon 3, University of Lyon, Institute of Business Administration, Lyon, France.*

Abstract. Project management is a field that has been applied in various areas of knowledge, particularly in engineering and software development. For organizations, projects are a central element for generating value. They allow to reach the organizational goal by using specific methodologies, tools and software. One of the most recognized tools, even in other fields of knowledge, for its impact on process improvement is maturity models. These models have already begun to be implemented in project management. Project Management Maturity Models are useful tools to evaluate the management process using a process reference (e.g., PMBOK). This process reference describes the best practices to achieve success in projects. The purpose of this paper is the identification of research papers that present maturity models specifically for project management. A useful classification for project managers using maturity models in a project management context is generated from the results of the review.

Keywords: project management; maturity models; systematic review.

For citation: Ruiz-Lopez J. F., Ortiz-Hernandez J., Bonjour E., Micaelli J.-P., Hernandez Y. Project Management Maturity Models: A Systematic Review. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 83-102. DOI: 10.15514/ISPRAS-2024-36(6)-5.

Модели зрелости управления проектами: систематический обзор литературы

[1] *Х. Ф. Руис-Лопес, ORCID: 0009-0003-9888-8537 <d19ce063@cenidet.tecnm.mx>*
[1] *Х. Ортис-Эрнандес, ORCID: 0000-0002-6481-1537 <javier.oh@cenidet.tecnm.mx>*
[2] *Э. Бонжур, ORCID: 0000-0001-7851-1095 <eric.bonjour@univ-lorraine.fr>*
[3] *Ж.-П. Микаэлли, ORCID: 0000-0002-9403-1419 <jean-pierre.micaelli@univ-lyon3.fr>*
[1] *Я. Эрнандес, ORCID: 0000-0002-8842-0899 <yasmin.hp@cenidet.tecnm.mx>*

[1] *Национальный технологический университет Мексики,*
Национальный центр исследований и технологических разработок, Куэрнавака, Мексика.
[2] *Исследовательская группа по инновационным процессам,*
Лотарингский университет, Нэнси, Франция.
[3] *Исследовательская лаборатория Магеллана, Лион 3,*
Лионский университет, Институт управления бизнесом, Лион, Франция.

Аннотация. Управлять выполнением проектов необходимо в различных областях знаний, особенно в области проектирования и разработки программного обеспечения. Для организаций проекты являются центральным элементом создания стоимости. Они позволяют достигать поставленных целей с помощью конкретных методологий, инструментов и программного обеспечения. Одним из наиболее признанных инструментов по степени влияния на улучшение процессов, и не только в этой области знаний, являются модели зрелости. Эти модели уже начали внедряться в управление проектами. Модели зрелости управления проектами являются полезными инструментами для оценки процесса управления с помощью референтных процессов (например, PMBOK). Референтные процессы описывают лучшие практики, позволившие достичь успеха в реализации проектов. Целью данного документа является выявление исследовательских работ, которые описывают модели зрелости, предложенные специально для управления проектами. На основе проведенного анализа для руководителей проектов, использующих модели зрелости в контексте управления проектами, создана полезная классификация.

Ключевые слова: управление проектом; модели зрелости; систематический обзор.

Для цитирования: Руис-Лопес Х. Ф., Ортис-Эрнандес Х., Бонжур Э., Микаэлли Ж.-П., Эрнандес Я. Модели зрелости управления проектами: систематический обзор литературы. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 83–102 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–5.

1. Introduction

Today, maturity models have been adopted by various industries and knowledge domains, including human resources, quality management, software development processes, manufacturing project management, products, and supply chain. This adaptation is due to the three main uses of maturity models reported in [1]: (1) to measure the level of maturity, (2) to provide a guide to achieve the maximum level of maturity, and (3) to establish a comparison with other organizations. In general terms, a maturity model can be defined as a collection of best practices that assist organizations in improving their processes[2]. According to [3], in 1986, the Software Engineering Institute began the development of a process maturity framework aimed at helping improve its software process. The first maturity model was published in 1988 [4] and called the Capability Maturity Model (CMM) by the Software Engineering Institute at Carnegie Mellon University in the United States, funded by the U.S. Department of Defense. The goal of this model was to assess the quality and capability of software companies providing services to the U.S. Department of Defense.

Due to its great utility, the models have been adopted by other domains such as processes, organizational management, software development, human resources, quality, project management, product development, and supply chain. In general, organizations recognize models as useful tools for assessing the status of specific or general processes. They allow for determining whether

processes have room for improvement. Maturity models present significant complexity because they lack a standardized manual for their application and may involve terminological biases that are not commonly used by those implementing them [5]. Models also require a considerable number of resources and time to be implemented properly and obtain a meaningful evaluation [6].

Garzás [7] stated that international efforts have been made to narrow down and adapt the software maturity models. Despite the difficulty of their implementation, software maturity models have given rise to project management maturity and assessment models. This difficulty lies in the fact that they are conceived as reference frameworks that establish the criteria for the operation of the area to be evaluated, based on pre-established manuals and/or standards, and do not explicitly state how to perform the evaluation. The most widely used manual or standard for building project management maturity models is the Project Management Book of Knowledge (PMBOK).

Many companies are motivated to implement project management to reduce process errors and drive improvements. In other cases, the motivation is to comply with standards or certifications that companies must meet to sell their products internationally. Based on this need to develop projects, companies use methodologies and tools, in some cases with computer support, to ensure their success.

The architecture of a system for predicting the performance of a project based on the evaluation of the project management maturity criteria is presented in [8]. In this research, an analysis of the mission statements of Sopra Steria Consulting and of the available literature on project management and maturity models has been carried out. It was found that, due to a lack of clarity in the concepts, the current maturity models are ambiguous in the way they should be applied in organizations. A standardization of some categories is proposed, which are included in a model called Invariant Based Maturity Model (IB2M). Also, a causal model is proposed to prove the existence of a relationship between project management maturity and cost overruns, showing that the maturity of the project management process is a significant determinant of the risk of cost overruns. This work is relevant to the research since it seeks to improve the area of project management using a tool such as the maturity assessment and for its methodological proposal to group and conceptualize in a clear and precise way maturity assessment criteria.

The remainder of the paper is organized as follows. Section 2 describes the systematic review process. Section 3 presents the analysis of the systematic review considering the quantitative and qualitative approaches. Section 4 describes the proposed classification of project management maturity models. In section 5, the future directions of project management maturity models are described. In section 6, a discussion is presented considering the results of this research. And finally, in section 7, the conclusions of this systematic review describe the main findings.

2. Systematic review methodology

The methodology of the systematic review is that proposed by [9]. It consists of three main phases: review planning, implementation of the review, and systematic reporting of the review. The detailed steps followed during the implementation of this state-of-the-art review methodology are described below.

2.1 Review planning

In this phase, the following aspects were identified and integrated as relevant for this precursor phase to the implementation of the review: rationale, approach, research question, and criteria.

2.1.1 Identification of the need for a systematic review

For the characterization of the justification, the particular interest in identifying the key aspects of the project management maturity models was considered to compare them and identify the areas of opportunity and the relevance of each model identified.

2.1.2 Development of the review protocol

A first version of the protocol was developed and analyzed to determine if it complied with two important aspects: completeness and consistency with the objectives of the review. The protocol was composed of the following parts: background, strategy for formulating the research question, strategy for selecting primary studies, selection criteria, strategy for establishing quality assessment criteria, data extraction strategy, synthesis strategy, dissemination strategy, and establishment of a schedule of activities.

2.2 Implementation of the review

In this phase, the systematic review protocol is used to develop the tasks established in the methodology for this phase. Previously the protocol went through a process of refinement and revision.

2.2.1 Identification of the research

In this section, the research question is formulated to help to identify project management maturity models. For the formulation of this question, the great importance of engineering project management maturity models was taken into consideration. Fig. describes the research question proposed.

What are the project management maturity models reported in the literature?

Fig. 1. Research Question.

2.2.2 Selection of primary studies

The selection followed the strategy outlined in the systematic review protocol, which consisted of searching for publications using search terms in search engines and repositories. Fig. 1 describes the research strings used in search engines in Spanish and English.

- *Improvement AND Project AND Management AND engineering,*
- *Evaluation AND Maturity AND Project AND Engineering,*
- *Evaluación AND Madurez AND Proyectos AND Ingeniería (research string used in Spanish),*
- *Maturity evaluation AND Project management,*
- *Mejora AND Proyecto AND Gestión AND Ingeniería (research string used in Spanish),*
- *Maturity evaluation AND project management AND SMES,*
- *Maturity evaluation and project management and systematic review.*

Fig. 1. Research Strings.

It was identified that the main keywords of the search string were as follows: Maturity and Project Management. A total of 1423 articles were reviewed considering only the title and keywords, after this review only 78 articles were considered primary studies. The inclusion and exclusion criteria in this phase of the process were only the title and keywords.*2.2.3 Characterization of the quality of publications.*

At this point, the following information was extracted for each publication: name of the article, number of pages, focus, number of citations, year of publication, and keywords, in addition to checking the availability of the article. After characterization, inclusion, and exclusion criteria were applied to each publication. The inclusion and exclusion criteria that were considered for the selection of articles to answer the research question are the following:

- Title: The title of the articles must include at least the keywords: Project Management and Maturity Model.
- Year of publication: Only articles from the last 10 years from 2012 to 2022 were selected.
- Number of citations: The number of citations of the articles was considered to determine their relevance to this research. For publications from 2022 and 2020, no filter applies, 2019 at least 4 citations, from 2018 at least 6 citations, from 2017 at least 8 citations, from 2016 at least 10 citations, from 2015 at least 12, from 2014 at least 14, from 2013 at least 16, from 2012 at least 18.
- Language: The languages selected were English and Spanish.

After the application of the inclusion and exclusion criteria only 32 articles were recognized as relevant to this research.

2.2.4 Data extraction and synthesis.

This step was carried out only based on the publications that passed the quality filters and were identified as relevant. In the synthesis task, each of the relevant publications was analyzed to identify project management maturity models, and a summary was produced because of this analysis.

2.3 Systematic reporting of the review

In this phase, a quantitative and qualitative analysis of the literature on project management maturity models was carried out. The objective of this analysis is to compare the publications and identify their contributions in terms of project management maturity models. Likewise, this stage seeks to show the results obtained from the implementation of the systematic review.

2.3.1 Quantitative analysis

The objective of the quantitative analysis provides a detailed overview of the evolution and distribution of relevant publications in the field of project management maturity models.

2.3.2 Qualitative analysis

The qualitative analysis provides an in-depth examination of relevant publications on project management maturity models, exploring various aspects and approaches.

3. Systematic review analysis

The existence of many articles related to project management maturity models involves a large amount of analysis time and depends on the subjectivity of the reader to determine their relevance. The main objective of systematic reviews is to apply a methodology to identify relevant papers reported in the literature that answer one or more research questions. To answer these questions, it is necessary to perform a qualitative and quantitative analysis of the articles and identify areas of opportunity not reported in the literature.

3.1 Quantitative analysis

The objective of the quantitative analysis is to provide a detailed overview of the evolution and distribution of relevant publications in the field of project management maturity models. This

includes analyzing the annual trend of publications, the distribution by search engines and repositories, identifying the most cited publications and the most frequent keywords, as well as evaluating the effectiveness of the search strings used. This approach highlights the importance and impact of these models in the scientific community and industry.

3.1.1 Publications by Year

With the quantitative analysis of this section, it was possible to identify that in 2014 there were a total of 7 relevant publications on project management maturity models, and in 2018 a total of 5 publications. In

Fig. 2, we observe this trend in the increase in the number of publications highlighting the importance of maturity models and project management for the scientific community and industry.

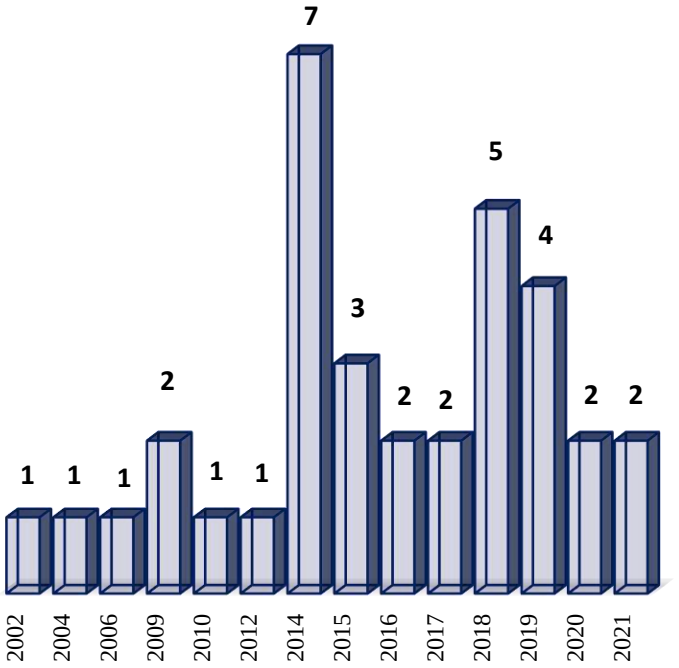


Fig. 2. Number of relevant PMMM publications by year.

3.1.2 Publications by Research Engine

To select the search engines, we considered the reports in the literature on digital libraries and indexing systems highly adopted by the scientific community. The most recognized search engines are Google Scholar and Worldwide Science, which is also highlighted in the scientific community for being the only search engine powered by Deep Technologies, a company dedicated to extracting scientific knowledge from the Deep Web.

In

Fig. 3, a pie chart shows the proportionality of the number of publications identified by title and keyword. In blue, the search engine Google Scholar with 61 publications; in orange and the search engine WorldWide Science with five publications.

3.1.3 Publications by Repository

The repositories consulted for this research were ScienceDirect and ACM Digital Library. However, other repositories were also identified where at least one publication on maturity models and project management was found.

Fig. 4 shows the number of publications identified by the repository in a bar chart. The repositories with the highest number of publications identified were Science Direct and Xplore.

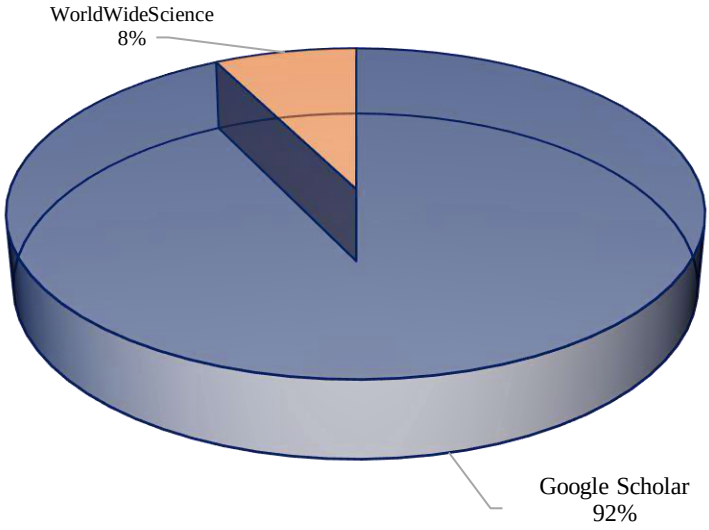


Fig. 3. Distribution of publications on PMMM by search engine.

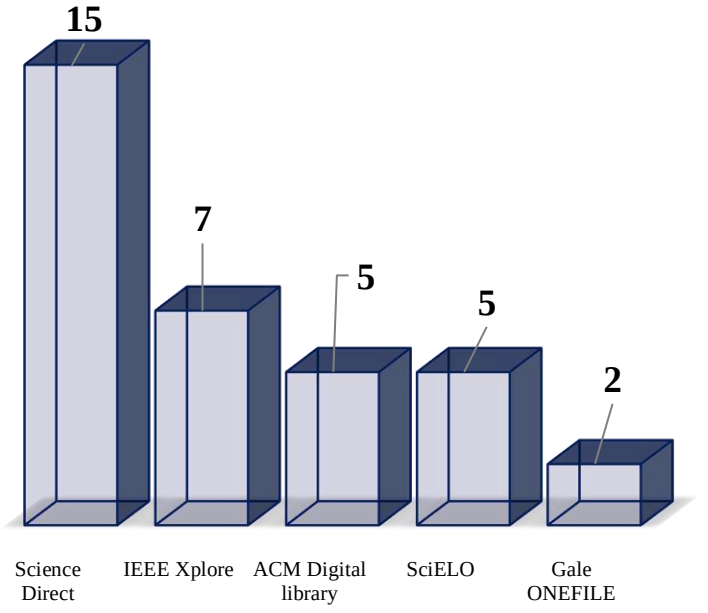


Fig. 4. Number of PMMM publications by repository.

3.1.4 Most Cited Publications

The collection of the number of citations made it possible to identify the most relevant publications on project management maturity models.

Fig. 5 describes the 10 most cited publications after applying the quality criteria to determine their relevance

3.1.5 Keywords

The analysis of the keywords of the publications on project management maturity models was useful in identifying current trends in this subject. Publications were identified that mixed the keywords: project management, maturity models, and software development, with which, it is easy to identify that at the beginning the maturity models were intended to evaluate in a general way to software development companies, which caused it to be implemented in project management of these types of companies. Consequently, due to their usefulness, the models began to be used in other fields of application such as engineering.

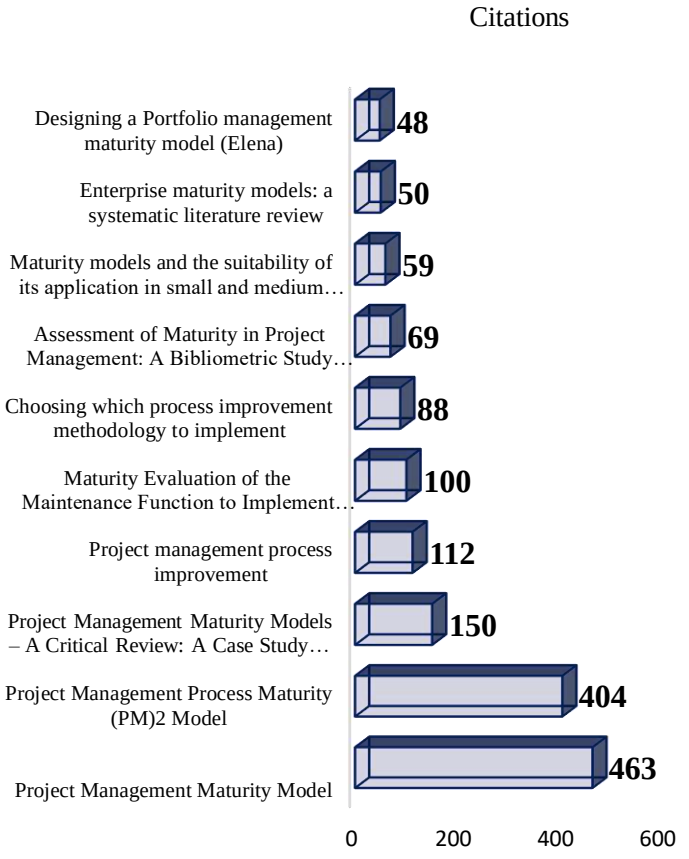


Fig. 5. Ranking of PMMM publications.

3.2 Qualitative analysis

The objective of the qualitative analysis is to provide an in-depth examination of relevant publications on project management maturity models, exploring various aspects and approaches. This includes analyzing keywords to identify trends and areas of focus, the limitation of maturity

model usage to large enterprises due to their resources and a detailed summary of studies and methodologies proposed in different publications. The studies review practical cases and adapted models to assess and improve maturity in different types of organizations, from large companies to SMEs, highlighting challenges, benefits, and recommendations for the effective implementation of these models.

3.2.1 Keywords of publications

Fig. 7 shows a cloud diagram of the keywords contained in the 32 articles identified as relevant. The words that were repeated the most are shown with a larger size, and the words that were not repeated as often, but were key, are shown with a smaller size.

Fig. 7. Keywords of publications.

3.2.2 General overview of project management maturity models

It was also identified that maturity models emerge as a proposal to improve the quality of software development because these models help to systematize processes and guarantee reproducibility within the application area. Subsequently, the models were applied to the areas that had a great impact on software development. In this case, project management was recognized as fundamental because commonly every development process is conceptualized as a project, which has a beginning, a process, and an end.

After reading and analyzing each of the publications, 12 project management maturity models were identified (OPM3, P3M3, CP3M, CIM3, PMMM, KPM3, MMGP, PM2, NPM3, PM2TOM2, P2MM, and P2CMM). Thanks to this approach of the maturity models, the project management area implemented as one of the useful tools to improve the process and identify good or bad practices that are performed within the area and determine the status concerning the ideal. The models use questionnaires as a method of evaluation to determine according to their results the level of maturity

of project management. In some cases, the models present 5 maturity levels and in others 4, however, in general, the models do not present a wide difference between the proposed levels.

The use of project management maturity models is mostly limited to large companies because they have financial and human resources that allow them to better evaluate their level of maturity and, for this reason, SMEs are limited in their use. The World Economic Forum has recognized that SMEs currently represent between 90 and 98% of the total number of companies in the world [10]. SMEs have an important role in the industry because many times these types of companies are suppliers from big companies and require a good project management maturity level.

4. Classification of PMMM

4.1 Comparison of PM maturity models

The qualitative and quantitative analysis of the articles on project management maturity models allowed us to answer the research question: What are the maturity models reported in the literature? It was also found that the maturity models have been extended to different research areas (processes, people, quality, manufacturing, products, supply chains, and operations). However, although a total of 12 PMMMs were identified, not all maturity models have the same assessment approach.

The complexity of visualizing these approaches is generated because the models proposed in the literature do not explicitly describe the evaluation process. With the systematic review, it was possible to identify articles that present practical cases that describe maturity evaluation processes in some companies, which helped to visualize the differences more explicitly. To visualize the differences between the maturity models identified, a comparison in Table 1 was built. The criteria used to compare the maturity models are:

- Approach
- Domain
- Maturity levels
- Number of best practices
- Evaluation process
- Number of survey questions
- Dimensions
- Referential
- Generated by the industry/scientific community

These criteria are important to identify the important information for our approach. In the beginning, it was considered the tools used to obtain the data but was identified that the survey was the only way to get the information on the process to be assessed. Also, it was identified that the maturity models are focused on traditional project management.

Table 1 describes the identified maturity models and makes a comparison. As a result of the comparative analysis, the findings for each criterion are described as follows.

4.1.1 Approach

The approaches identified in the literature for project management maturity models are project management, the organizational performance of the project management area, construction project management, and the project management process in companies. These different approaches show that the companies recognize the process of management of the area, the role of the project management area in the company, and the project management as different entities that need a specific project management maturity model.

Table 1. Project Management Maturity Models reported in the literature.

Model	Approach	Domains	Maturity levels	Number of best practices	Evaluation process	Number of questions in the survey	Dimensions	Based on	Generated by
OPM3: Project Management Maturity Model	Project Management	Project Management Program Management Portfolio Management	Level 1: Ignorance Level 2: Medium Level 3: Maximum	600	Step 1: Preparation for the evaluation Step 2: Assessment Step 3: Improvement plan Step 4: Implement improvement Step 5: Repeat the process	150 (Self-assessment)	1.- Domains 2.- Process improvement steps	PMBOK	Industry
CP3M V5.0: Colombian Project Management Maturity Model	Project Management	Project Management Program Management Portfolio Management	Level 1: Inconsistency Level 2: Planning and Control Level 3: Integration Level 4: Strategic alignment Level 5: Innovation and optimization	Not identified	Step 1: Process Inventory Step 2: Evaluation scheduling Step 3: Organization/ Project Characterization Step 4: Assessment of the level of formalization of practices Step 5: Information processing Step 6: Definition of process capability level Step 7: Maturity level definition Step 8: Analysis and presentation of results	The evaluation method is based on the criteria established for each maturity level	1.- PMBOK 2.- Strategic alignment 3.- Learning 4.- Adaptability 5.- Life Cycle	PMBOK	Scientific Community
MMGP: Project Management Maturity Model- Darci Prado	Project Management	Project Management	Level 1: Initial Level 2: Knowledge Level 3: Standardized Level 4: Managed Level 5: Optimized	Not identified	Not identified	40	1.- Strategic alignment 2.- Behavioral competence Organizational structure 4.- Computerization 5.- Methodology 6.- Technical competence	PMBOK	Scientific Community
NPM3: National Project Management Maturity Model	Project Management	Project Management	Level 1: Emerging Level 2: Developing Level 3: Adolescent Level 4: Maturity	Not identified	Not identified	The number of questions for the maturity assessment is not explicitly described	Not identified	Not identified	Scientific Community
PM2TOM2: Project management methods and tools-oriented maturity model	Project Management	Project Management	1.- Low Management Maturity Level 2.- Lower Medium Management Maturity Level 3.- Medium Management Maturity Level 4.- Advanced Management Maturity Level 5.- High Management Maturity Level	Not used, the model uses 43 methods and tools to assess	Step 1: Collection of data in the organization carry out the projects Step 2: Analysis of the data obtained using the model designed Step 3: Evaluation of project management maturity Step 4: Use of the evaluation results	The number of questions for the maturity assessment is not explicitly described	1.- Time management 2.- Resource management 3.- Cost management 4.- Risk management 5.- Scope Management 6.- Organizational support of the project 7.-Staff training, and project management support based on literature review	Based on the literature review	Scientific Community
P2MM: Prince 2 maturity model	Project Management	Project Management Program Management	1.- Awareness of the process 2.- Repeatable process 3.- Defined process	Not identified	Not identified	The number of questions for the maturity assessment	1.- Management Control 2.- Benefits Management	Prince 2	Scientific Community

		t Portfolio Management	4.- Managed process 5.- Optimized process			is not explicitly described.	3.- Financial Management 4.- Stakeholder Engagement 5.- Risk Management 6.- Organizational Governance 7.- Resource Management		
P2CMM: Portfolio Management Maturity Model	Project Management	Project Management	1.- Cognitive level 2.- Repeatable level 3.- Management level 4.- Integration level 5.- Continuous level	This model merges 45 sub-processes of PRINCE2 into 25 sub-processes	The authors describe a process to use the data after the evaluation and describe the type of questionnaire used	The number of questions for the maturity assessment is not explicitly described.	1.-Starting Up a project 2.- Initiation a project 3.- Directing a project 4.- Controlling a stage 5.- Managing product delivery 6.-Managing stage boundaries 7.- Closing a project 8.- Planning	Prince 2	Government
P3M3: Program and Project Management Maturity Model	Organizational performance of the project management area	Project Management Program Management Portfolio Management	Level 1: process knowledge Level 2: a repeatable process Level 3: defined Process Level 4: managed process Level 5: optimized process	Not identified	Step 1: Planning the assessment Step 2: Selection of the P3M3 model to be used Step 3: Establishing the scope Step 4: Selecting the data structure Step 5: Selecting the approach Step 6: Planning the checklist Step 7: Understanding results Step 8: Improving planning	9 (Self-assessment)	Not identified	Not identified	Industry
PMMM: PM Solutions Project Management Maturity Model	Project management process in companies	Project Management Program Management Portfolio Management	Level 1: Initial process Level 2: Structured and standardized process Level 3: Organizational standards and institutionalized processes Level 4: Process managed Level 5: Process optimized	Not identified	Step 1: Analyze the status of the company Step 2: Characterization of the company's practices Step 3: Analysis of surveys built based on PMMM Step 4: Determination of maturity level	52	1.- Scope management 2.- Time management 3.- Cost management	PMBOK	Industry / Scientific Community
KPM3: Kerzner Project Management Maturity Model	Project management process in companies	Project Management Program Management Portfolio Management	Level 1: Common language Level 2: Common process Level 3: Unique methodology Level 4: Benchmarking Level 5: Continuous Improvement	Not identified	Step 1: Analyze the current status of the company Step 2: Characterization of the company's practices Step 3: Analysis of surveys built based on PMMM Step 4: Determination of maturity level	183	Not identified	CMMM	Scientific Community
PM2: Project Management Process Maturity Model	Project management process in companies	Project Management	Level 1: Basic Project Management Process Level 2: Individual project planning Level 3: Systematic project planning and control Level 4: Multi-project integration planning and control	Not identified	Not identified	148	Not identified	Not identified	Scientific Community

			Level 5: Continuous improvement of the project management process						
CIM3: Construction Industry Macro Maturity Model	Construction project management	Project Management	Level 1: Immaturity (0) Level 2: Immaturity (1/3) Level 3: Traditional maturity (2/3) Level 4: Maturity (1)	They propose key practices which are generated with the objective of achieving the organizational objective. For this reason, maturity levels are established based on the existence of key practices	Step 1: Establishing the importance of key practice areas Step 2: Organizational capability	The evaluation method considers the assessment of the existence of key practices	1.- Cost 2.- Quality 3.- Health and safety 4.- Human resources	CMMM	Industry

4.1.2 Domains

The different domains identified are project management, program management, and portfolio management. According to [35], Project management is defined as the application of knowledge, abilities, and tools to project activities to accomplish the requirements. Program management also is defined as the application of knowledge, abilities, and tools to a program's activities to complete it and get benefits by managing program components together. Finally, Portfolio management is defined as the central management of one or more portfolios to achieve strategic objectives. Considering these definitions provided by the PMBOK, the different domains of project management are considered by the scientific community and industry to create the models.

4.1.3 Maturity Levels

Project management maturity levels are considered as the steps that organizations must climb to position themselves in terms of project management practices. This positioning allows for identifying the status and the improvement plan needed to reach higher levels. In general, maturity models contemplate 5 maturity levels. From the comparison, it was identified that 75% of the identified models present 5 maturity levels, 17% present 4 maturity levels, and 8% present 3 maturity levels.

4.1.4 Number of best practices

The practices of project management are the number of activities related and aligned to achieve the project objectives. Some of the maturity models consider the best practices described by manuals or referential to establish a maturity level and to create surveys to assess maturity.

Considering the Project management maturity models identified in this research, 8% of the models describe the number of best practices considered, 67% of the models do not describe the best practices considered, and 25% of the models describe an alternative criterion considered (processes, method, and tools).

4.1.5 Evaluation process

The evaluation process is the description of the steps required to assess the enterprises. Normally, the Project management maturity models describe the criterion that needs to be considered in the assessment, and many times the survey is proposed. However, the vast majority (proposed by industry) do not explicitly describe the process that must be followed to perform the maturity assessment. In this research, the evaluation process was identified in the articles that apply a specific maturity model in a case study.

4.1.6 Number of questions in the survey

The surveys have an important role in the project management maturity models, due to this being the principal tool to obtain data from the enterprises. Some of the maturity models offer a list of questions that are used to apply the questionaries and make a self-evaluation of the maturity. The self-evaluation is the action of applying a questionary in the enterprise without contracting an external to perform the maturity assessment. This term helps to identify that there are two different ways to perform a project maturity assessment, on one side the self-evaluation and on the other the external evaluation. In Table 1, 50% of the models identified describe the number of questions and the other 50% of the models do not describe the number of questions and neither the content of the questions.

4.1.7 Dimensions (Knowledge areas)

The knowledge areas or dimensions are defined in [35] as an area identified in project management that is determined by the knowledge requirements and described in terms of its component process, practices, inputs, outputs, tools, and techniques. The maturity models consider the dimensions to select what areas of project management going to evaluate and generate the questions. Considering the comparison in Table 1, the maturity models describe at list the time, resource, scope, quality, and risk as principal dimensions considered for the assessment. However, there are some other dimensions mentioned such as organizational governance, financial management, staff training, and others.

4.1.8 Referential

The PMBOK, Prince 2, CMMM, and literature review are the main referential considered by the Project management maturity models identified in Table 1. For the maturity models to be relevant the selection of the referential to generate the model due to here is described the best practices, domains, and approaches. These references provide a strong scientific base for the maturity model but many times this is a limitation when a new version drops.

4.1.9 Generated by the industry or the scientific community

It was identified that the scientific community and the industry are the main suppliers of the project management maturity models. Also, the government identified the proposed maturity model. This helps to understand why many Project management maturity models describe with more detail the steps to assess the maturity, the questions, and also the requirements. The models that are proposed by the industry don't explain explicitly the evaluation process due to they look to contact the enterprises to perform this assessment. The models proposed by the scientific community are focused in proportionate the major quantity possible of information for the people that implement the proposal. This means that there are two main purposes for generating a model: the consulting business and the academic validation.

4.2 Suggested Classification

The different approaches make it possible to generate a suggested classification that groups the different models according to their objectives. This section shows the classification proposed after analyzing the publications selected as relevant to the systematic review of project management maturity models. Fig. 8 presents the classification of project management maturity models according to their approach.

4.2.1 Project Management

The models in this category are focused on measuring the maturity of the projects. In this case, the models in this category have the following objectives:

- OPM3: This standard defines a methodology for implementing and improving OPM. It comprises a five-step iterative cycle that emphasizes evaluation and continuous improvement. In the broadest sense, OPM3® is a maturity model.
- CP3M V5.0: CP3M© has been established as a formal instrument to measure the maturity of an organization's project management.
- MMGP: The MMGP model was created to help the project management team of the Instituto de Desenvolvimento Gerencial (INDG), currently Falconi Consultores de Resultado, in the evaluation of the maturity status of the organizations that hire it.
- NPM3: The NPM3 model was developed to increase the maturity of organizational project management in national contexts.
- PM2TOM2: The PM2TOM2 model was created to evaluate the project management maturity based on the assessment of the usage of project management tools and methods in each stage of the project life cycle.
- P2MM: the P2MM model was created to provide a framework to evaluate the actual adoption of the method PRINCE2 and provide improvement plans based on industry best practices.
- P2CMM: The P2CMM model considers the PRINCE2 approach to make a qualitative evaluation of the process of project management and operation considering an evaluation index system.

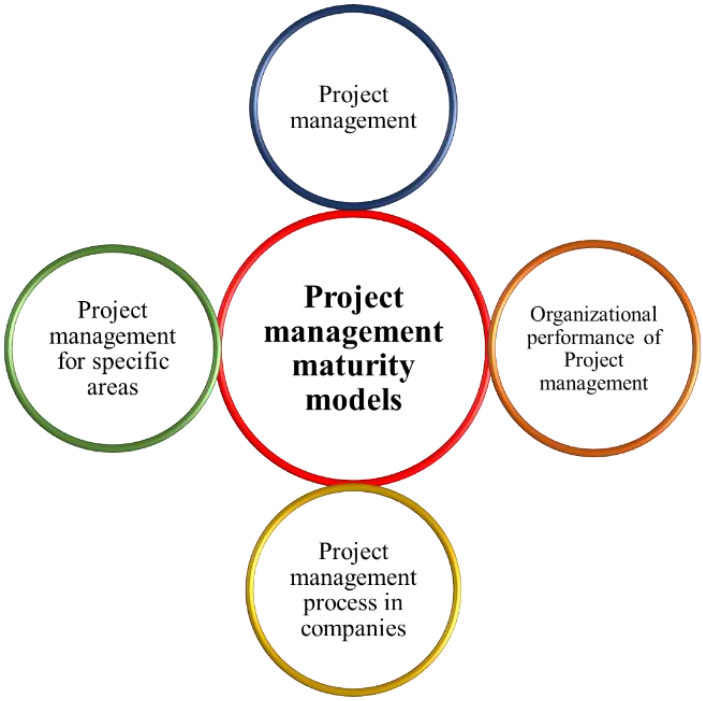


Fig. 8. Classification of project management maturity models.

4.2.2 Organizational project management performance

The model identified in this category is focused on evaluating the organizational performance of the project management area. That is, it specifically seeks to determine how well or poorly the project management area interacts with the impacted areas. The identified model has the following objective:

- P3M3: Framework to evaluate and benchmark your organization's current performance and develop improvement plans.

4.2.3 Project management process in companies

The models that fall into this category are focused on evaluating the project management process in companies to determine if the project management process is performed according to the best practices described in each model. The models in this category have the following general objectives.

- PMMM: The PMMM model aims to develop organizational performance within companies. This model uses the areas of knowledge proposed by the PMI to have more specific measures.
- KPM3: The KPM3 was created for developing the organizational capabilities and culture to incorporate project management practices into the organization's processes and procedures.
- PM2: Integrate previous project management practices, processes, and maturity models to improve the effectiveness of project management in the organization. organization.

4.2.4 Area-specific project management models

In this category, some models were generated for a specific area related to project management. In such a case, the model identified has the following objective:

- CIM3: The objectives of CIM3 are to model the construction industry maturity at the macro level to provide project performance indicators; to provide a context in which to interpret project performance; to allow comparisons between various regions, and to provide guidance on the construction industry performance improvement initiatives.

5. Future directions

Project management maturity models, despite their great usefulness in the conduct of assessments, have some important limitations in their use for organizations that do not have the necessary resources or organizational structure to implement them. In addition to these limitations, organizations face a difficult challenge in determining how to perform the assessments without a step-by-step guide describing how to perform the assessment process.

The generation of project management maturity model ontologies can help provide a conceptual understanding of the assessment process and even the relevance of the assessment. Although project management maturity models are implemented using surveys as a tool to extract information from the process, they require a trained staff and a group of people in charge of applying them. This tool is effective for companies that have the budget for continuous improvement. When this is not the case, it becomes a limitation in its applicability. For this reason, it is considered necessary to generate another tool to extract information from the project management process that does not require highly trained personnel or many resources to extract information from the process.

Likewise, it has been observed that some project management maturity models generated by the industry do not detail the maturity assessment process; they only limit themselves to explaining what the assessment criteria are, in this case, the good project management practices that should be carried out. The literature describes some case studies of some maturity models that could help implement maturity assessments.

6. Discussion

The interest of this systematic review was to identify the project management maturity models reported in the literature. To ensure the reproducibility of the present research, the systematic review

methodology of [4] was selected and the steps to perform the systematic review focused on identifying project management maturity models were described.

In this review, articles published between 2020-2021 were considered and the search engines Google Scholar and Worldwide Science were used. The repositories Science Direct, IEEE Xplore, ACM Digital Library, and SciElo, among others, were also considered. The quality criteria considered were article title, keywords, number of citations, year of publication, language, and availability. As a result, 26 maturity models were identified, focused on processes, software, human resources, quality, project management, manufacturing, products, and supply chain. Of the total number of models identified, only 12 are project management, maturity models. The identified models were grouped according to their objective: a) project management (OPM3, CP3M V5.0, MMGP, NPM3, PM2TOM2, P2MM, and P2CMM), b) organizational performance of project management (P3M3), c) project management process in companies (PMMM, KPM3, PM2) and d) project management models for specific areas (CIM3).

However, another review of the state of the art published in 2022 [36], was identified in the literature, focused on providing recommendations through the analysis of project management maturity models to provide recommendations for selecting or generating a project management maturity model. This review took into account articles published up to 2022, retrieved from the Scopus and Web of Science repositories. The inclusion and exclusion criteria for selecting relevant articles were rank, article citation index, and application rank. The authors propose to group the maturity models into three groups: a) Maturity Models of leading PM Organizations (OPM3, IPMA Delta Model, P3M3, and P2MM), b) Most cited and validated historical Maturity Models (CMMI, PM2, KPMMM, PMMM, and PMM), c) More recent maturity models (NPM3, MMM, and SPM3).

To compare this work with the work of [36], we found that the proposed classification of project management maturity models focuses on grouping maturity models according to their relevance and complexity. This classification is interesting if the objective is to show the most relevant project management maturity models considering specific criteria such as complexity, year of publication, organization, maturity levels, and domain.

Instead, our proposal is focused on grouping maturity models considering criteria such as approach, domain, maturity levels, number of best practices, evaluation process, number of survey questions, dimensions, and referential, generated by the industry/ scientific community. These criteria were selected because the main challenge that enterprises face when implementing a maturity model is the evaluation process. This process is not described in an explicit way in the literature. One of the aims of this systematic review was to identify the evaluation process for each model to provide the steps described to implement the maturity model of project management in the enterprise.

Normally the maturity models are implemented by project managers and personnel involved in the process of project management. For this reason, providing a complexity evaluation before the managers consider their available resources and structure is not viable. Each enterprise has different needs, and several resources available destined to improve the process, and size.

Project management maturity models use surveys as an assessment tool to extract information. This tool involves the use of a large amount of financial and human resources, as well as a full understanding of the project management process. If companies require resources and knowledge to implement it, will maturity models be applicable to all types of companies? To answer this question, we will discuss the challenges faced by SMEs to implement the models: 1) Limited financial resources, 2) Untrained personnel, 3) Difficulty in using surveys as an assessment method, 4) Ambiguity in the assessment process, 5) Difficulty in choosing the most appropriate model.

The use of manuals or references (i.e., PMBOK) for generating project management maturity models tends to lose validity when a new version is generated. Therefore, the generation of project management maturity models from the analysis of the literature in which current and future trends on project management are considered can increase the time of use of these models.

7. Conclusions

This article presents the origins of maturity models, the models reported in the literature, and the maturity models of the specific area of project management, which in turn have different approaches despite being in the same area.

The systematic review involved a total of 4 repositories and 2 search engines in which a total of 1423 articles were reviewed by title and keywords. After applying the inclusion and exclusion criteria, only 32 articles were identified as relevant. As a result of the systematic review and analysis of the articles, a classification according to their focus was proposed: maturity models focused on project management (7), maturity models focused on the organizational performance of project management (1), project management process maturity models (3) and finally, project management maturity models specifically designed to evaluate a single type of project (1).

The importance of maturity models is increased by the fact that nowadays all companies are trying to be more competitive and gain an advantage, which pushes them to continuously adopt new tools and technologies that help them to improve their processes and make them more effective and efficient. For future work, it is recommended to conduct a comparative analysis of the project management maturity models of each category to determine which are more efficient according to each approach. This will help academics and project managers select the most appropriate maturity model for their organization.

References

[1]. L. Torres, “A contingency view on the effect of project management maturity on perceived performance,” Ph.D Dissertation, Business School SKEMA, Lile, 2014.

[2]. J. Garzás, C. M. Fernández, and M. Piattini, “Una aplicación de la norma ISO/IEC 15504 para la evaluación por niveles de madurez de Pymes y pequeños equipos de desarrollo”, Revista Española de Innovación, Calidad e Ingeniería del Software, vol. 5, no. 2, pp. 88-98, Jul. 2009.

[3]. Felipe Sánchez-Garzón, “Supporting the transformation of a company’s project management by elaborating an invariant-based project management maturity model and a causal predictive model between maturity criteria and project performance”, P.H.D Dissertation, Lorraine Université, Nancy, 2019.

[4]. B. Kitchenham and S. Charters, “Guidelines for performing Systematic Literature Reviews in Software Engineering”, 2007.

[5]. A. González, A. Sandoval, and B. Heredia, “Elaboración de planes de entrenamiento cruzado a personal clave en proyectos: el caso de una empresa de ingeniería”, Revista ingeniería de construcción, vol. 33, no. 3, Dec. 2018, doi:10.4067/S0718-50732018000300205.

[6]. Á. Cuadros López, J. A. Morales Viveros, and A.B. Rojas Melendez, “Propuesta metodológica para medir el nivel de madurez de la gestión de proyectos en empresas de ingeniería”, Revista EIA, vol. 14, no. 27, Sep. 2017, doi: 10.24050/reia.v14i27.808.

[7]. E. Pérez-Mergarejo, I. Pérez-Vergara, and Y. Rodríguez-Ruíz, “Modelos de madurez y su idoneidad para aplicar en pequeñas y medianas empresas”, Ingeniería Industrial, vol. XXXV, no. 2, pp. 146–158, 2014, [Online]. Available: https://www.redalyc.org/articulo.oa?id=36043 3597004.

[8]. F. F. Espinosa and G. E. Salinas, “Evaluación de la Madurez de la Función Mantenimiento para Implementar Innovaciones en su Gestión”, Información tecnológica, vol. 21, no. 3, 2010, doi: 10.4067/S0718-07642010000300002.

[9]. M. V. Jacquez-Hernández and V. G. López Torre, “Modelos de evaluación de la madurez y preparación hacia la Industria 4.0: una revisión de literatura”, Ingeniería Industrial. Actualidad y Nuevas Tendencias, vol. VI, no. 20, pp. 61– 78, 2018, [Online]. Available: https://www.redalyc.org/articulo.oa?id=21505 7003004.

[10]. Prado Palacio and Leidy Johanna, “Grado de madurez en gestión de proyectos de una empresa constructora de vivienda en Santiago de Cali: caso de Buenavista constructora y promotora S.A.S”, Revista Espacios, vol. 40, no. 43, pp. 20–33, 2019, [Online]. Available: https://bibliotecadigital.univalle.edu.co/handle/ 10893/15226

[11]. Ó. Rubiano Ovalle and A. Cuadros Lopez, “Estado de la Gerencia De Proyectos En Pymes Del Sector Artes Gráficas y Propuestas de Lineamientos para su Mejoramiento”, (State of Project Management in SMEs of the Printing Sector and Proposed Guidelines for Improvement), 2012.

[12]. J. Shan and R. Lu, “Research on Application of Project Management Maturity Model in Risk Management of Electrical Engineering”, E3S Web of Conferences, vol. 218, Dec. 2020, doi: 10.1051/e3sconf/202021802040.
[13]. D. Yao, R. Zhou, and T. Fan, “Research on maturity of engineering project informatization management—As an example of Anhui Sanjian Engineering Co., LTD”, IOP Conf Ser Earth Environ Sci, vol. 267, p. 042144, 2019, doi: 10.1088/1755-1315/267/4/042144.
[14]. R. Morlhon, R. Pellerin, and M. Bourgault,“Building Information Modeling Implementation through Maturity Evaluation and Critical Success Factors Management”, Procedia Technology, vol. 16, 2014, doi: 10.1016/j.protcy.2014.10.127.
[15]. T.-C. Lin and K. J. Wang, “Project-based maturity assessment model for smart transformation in Taiwanese enterprises”, PLoS One, vol. 16, no. 7, Jul. 2021, doi: 10.1371/journal.pone.0254522.
[16]. F. Sanchez, E. Bonjour, J.-P. Micaelli, and D. Monticolo, “An Approach Based on Bayesian Network for Improving Project Management Maturity: An Application to Reduce Cost Overrun Risks in Engineering Projects”, Comput Ind, vol. 119, Aug. 2020, doi: 10.1016/j.compind.2020.103227.
[17]. F. Backlund, D. Chronéer, and E. Sundqvist, “Project Management Maturity Models – A Critical Review”, Procedia Soc Behav Sci, vol. 119, pp. 837–846, Mar. 2014, doi: 10.1016/j.sbspro.2014.03.094.
[18]. H. Tahri and O. Drissi-Kaitouni, “New Design for Calculating Project Management Maturity (PMM)”, Procedia Soc Behav Sci, vol. 181, pp. 171–177, May 2015, doi: 10.1016/j.sbspro.2015.04.878.
[19]. A. J. Christoph and S. Konrad, “Project Complexity as an Influence Factor on the Balance of Costs and Benefits in Project Management Maturity Modeling”, Procedia Soc Behav Sci, vol. 119, pp. 162–171, Mar. 2014, doi: 10.1016/j.sbspro.2014.03.020.
[20]. A. Miklosik, “Improving Project Management Performance through Capability Maturity Measurement”, Procedia Economics and Finance, vol. 30, pp. 522–530, 2015, doi: 10.1016/S2212-5671(15)01264-2.
[21]. Y. H. Kwak and C. W. Ibbs, “Project Management Process Maturity (PM)2 Model”, Journal of Management in Engineering, vol. 18, no. 3, pp. 150–155, Jul. 2002, doi:10.1061/(ASCE)0742-597X (2002)18:3(150).
[22]. A.-L. Mesquida and A. Mas, “A project management improvement program according to ISO/IEC 29110 and PMBOK®”, Journal of Software: Evolution and Process, vol. 26, no. 9, pp. 846–854, Sep. 2014, doi: 10.1002/smr.1665.
[23]. M. Gershon, “Choosing Which Process Improvement Methodology to Implement”, The Journal of Applied Business and Economics, vol. 10, no. 5, p. 61, 2010.
[24]. Robert K. Wysocki, Project Management Process Improvement. Norwood: ARTECH HOUSE, INC, 2004.
[25]. H. K. Mittermaier and H. Steyn, “Project management maturity: An assessment of maturity for developing pilot plants”, South African Journal of Industrial Engineering, vol. 20, no. 1, pp. 95–107, 2009.
[26]. T. F. de Souza and C. F. S. Gomes, “Assessment of Maturity in Project Management: A Bibliometric Study of Main Models”, Procedia Comput Sci, vol. 55, pp. 92–101,2015, doi: 10.1016/j.procs.2015.07. 012.
[27]. J. C. Viana and C. M. de M. Mota, “Enhancing Organizational Project Management Maturity: a framework based on the value focused thinking model”, Production, vol. 26, no. 2, pp. 313–329, Nov. 2015, doi: 10.1590/0103- 6513.1699 13.
[28]. S. Nikkhou, K. Taghizadeh, and S. Hajiyakhchali, “Designing a Portfolio Management Maturity Model (Elena)”, Procedia Soc Behav Sci, vol. 226, pp. 318–325, Jul. 2016, doi: 10.1016/j.sbspro.2016. 06.194.
[29]. N. Williams, N. P. Ferdinand, and R. Croft, “Project management maturity in the age of big data”, International Journal of Managing Projects in Business, vol. 7, no. 2, pp. 311–317, Apr. 2014, doi: 10.11 08/IJMPB-01-2014-0001.
[30]. F. Y. Hernández, R. I. Laguado, and J. P. Rodriguez, “Maturity analysis in project management in Colombian universities”, J Phys Conf Ser, vol. 1126, p. 012055, Nov. 2018, doi: 10.1088/1742-6596/1126/1/012 055.
[31]. D. Seelhofer and C. O. Graf, “National project management maturity: A conceptual framework”, Central European Business Review, vol. 7, no. 2, p. 1, 2018.
[32]. L. De Souza Scotelano, R. Dalvo Pereira da Conceição, U. Da Costa Leonídio, and C. Silva de Jesus, “Project management maturity model: the case in an automotive industry in Brazil”, Brazilian Journal of Operations & Production Management, vol. 14, no. 4, pp. 500–507, Dec.2017, doi: 10.14488/BJOPM.2017.v14. n4.a6.
[33]. K. Jana and T. Libena, “Proposal and Verification of Project Management Methods and Tools Oriented Maturity Model”, Journal of Management and Projects, vol. 9, no. 1, 2018.

[34]. R. Silva, N. Duarte, T. Barros, and G. Fernandes, "Project Management Maturity: Case study analysis using OPM3® model in manufacturing industry", in 2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), IEEE, Jun. 2019, pp. 1–8. doi: 10.1109/ICE.2019.8792586.

[35]. I. Project Management Institute, A Guide to the Project Management Body of Knowledge (PMBOK® Guide), 7th ed. Newton Square, Pennsylvania: Project Management Institute, Inc., 2021.

[36]. E. Fabbro and S. Tonchia, "Project Management Maturity Models: Literature Review and New Developments", The Journal of Modern Project Management, vol. 8, no. 3, pp. 34–45, May 2022.

Информация об авторах / Information about authors

Хорхе Франсиско РУИС-ЛОПЕС учится в аспирантуре факультета программирования Национального центра исследований и технологических разработок (Мексика).

Jorge Francisco RUIZ-LOPEZ – Doctoral student in Computer Science at the Computer Science Department of the National Center of Research and Technological Development in Mexico.

Хавьер ОРТИС-ЭРНАНДЕС имеет степень PhD по автоматизации и программированию, штатный профессор факультета программирования Национального центра исследований и технологических разработок (Мексика). Сфера научных интересов: цифровая трансформация, инженерия требований, моделирование бизнес-процессов.

Javier ORTIZ-HERNANDEZ – PhD in Automation and Computer Science, full professor at the Computer Science Department of the National Center of Research and Technological Development in Mexico. Research interests: digital transformation, requirements engineering, and business process modeling.

Эрик БОНЖУР – имеет степень PhD по автоматизации и промышленной инженерии, профессор. Специализируется в области системной инженерии в Национальной высшей школе инноваций и системного инжиниринга (ENSGSI, Лотарингский университет). Сфера научных интересов: системная инженерия, основанная на моделях, приемлемость инновационных решений, цифровая трансформация, управление знаниями и навыками.

Eric BONJOUR – PhD in Automation and Industrial Engineering, full Professor specialized in systems engineering at the National Graduate School of Innovation and Systems Engineering (ENSGSI, Université de Lorraine). Research interests: model-based systems engineering, acceptability of innovative solutions, digital transformation, knowledge and skills management.

Жан-Пьер МИКАЭЛЛИ имеет степень PhD по экономике промышленности от Университета Люмьер, Лион. Магистр и старший лектор по индустриальному и операционному управлению в Школе управления Лионского института управления бизнесом (Университет Жана Мулена). Научные интересы: разработка методов оценки производительности и инструментария для архитектурно сложных технических систем.

Jean-Pierre MICAELLI – PhD in Industrial Economics, Lumière University, Lyon, Master in Industrial Economics Jean Moulin University, Lyon, Senior Lecturer in Industrial and Operations Management at IAE School of Management of Lyon (Jean Moulin University), Research interests: development of performance evaluation methods and tools for the architecture of complex technical systems.

Ясмин ЭРНАНДЕС имеет степень PhD по программированию, профессор факультета программирования Национального центра исследований и технологических разработок (Мексика). Научные интересы: интеллектуальные обучающие системы, человек-машинное взаимодействие, интеллектуальный анализ данных в образовании, аффективные вычисления, машинное обучение.

Yasmín HERNÁNDEZ – PhD in Computer Science, full professor at the Computer Science Department of the National Center of Research and Technological Development in Mexico. Research interests: intelligent tutoring systems, human-computer interaction, educational data mining, affective computing, machine learning.

DOI: 10.15514/ISPRAS-2024-36(6)-6

Could an LLM Like chatGPT Perform a Functional Size Measurement using the COSMIC Method?

[1] *F. Valdés-Souto, ORCID: 0000-0001-6736-0666 <fvaldes@ciencias.unam.mx>*
[2] *D. Torres-Robledo, ORCID: 0009-0002-7168-9709 <dtorres@ciencias.unam.mx>*

[1] *National Autonomous University of Mexico, Science Faculty, CDMX, Mexico*.
[2] *National Autonomous University of Mexico*
Research Institute in Applied Mathematics and Systems, CDMX, Mexico.

Abstract. The process of developing software is intricate and time-consuming. Resource estimation is one of the most important responsibilities in software development. Since it is currently the only acceptable metric, the functional size of the program is used to generate estimating models in a widely accepted manner. On the other hand, functional size measurement takes time. The use of artificial intelligence (AI) to automate certain software development jobs has gained popularity in recent years. Software functional sizing and estimation is one area where artificial intelligence may be used. In this study, we investigate how to apply the concepts and guidelines of the COSMIC method to measurements using ChatGPT 4o, a large language model (LLM). To determine whether ChatGPT can perform COSMIC measurements, we discovered that ChatGPT could not reliably produce accurate findings. The primary shortcomings found in ChatGPT include its incapacity to accurately extract data movements, data groups, and functional users from the text. Because of this, ChatGPT's measurements fall short of two essential requirements for measurement: accuracy and reproducibility.

Keywords: COSMIC; CFP; functional size measurement; LLM; chatGPT; software engineering; AI; automatization.

For citation: Valdés-Souto F., Torres-Robledo D. Could an LLM like chatGPT perform a functional size measurement using the COSMIC method? Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024, pp. 103-114. DOI: 10.15514/ISPRAS-2024-36(6)-6.

Может ли языковая модель на базе chatGPT измерять функциональный размер методом COSMIC?

[1] *Ф. Вальдес-Соуто, ORCID: 0000-0001-6736-0666 <fvaldes@ciencias.unam.mx>*
[2] *Д. Торрес-Робледо, ORCID: 0009-0002-7168-9709 <dtorres@ciencias.unam.mx>*
[1] *Национальный автономный университет Мексики, факультет Науки, Мехико, Мексика.*
[2] *Национальный автономный университет Мексики, Исследовательский институт прикладной математики и систем, Мехико, Мексика.*

Аннотация. Процесс разработки программного обеспечения является сложным и трудоемким. Оценка ресурсов является одной из наиболее важных обязанностей в разработке программного обеспечения. Поскольку в настоящее время это единственный приемлемый показатель, функциональный размер программы используется для генерации моделей оценки общепринятым способом. С другой стороны, измерение функционального размера требует времени. Использование искусственного интеллекта (ИИ) для автоматизации определенных рабочих мест разработчиков программного обеспечения набрало популярность в последние годы. Определение размеров и оценка функциональности программного обеспечения является одной из областей, в которой может использоваться искусственный интеллект. В этом исследовании мы исследуем, как применять концепции и рекомендации метода COSMIC к измерениям с использованием ChatGPT 4o, большой языковой модели (LLM). Чтобы определить, может ли ChatGPT выполнять измерения COSMIC, мы обнаружили, что ChatGPT не может надежно производить точные результаты. К основным недостаткам, обнаруженным в ChatGPT, относится его неспособность точно извлекать из текста движения данных, группы данных и функциональных пользователей. Из-за этого измерения ChatGPT не соответствуют двум основным требованиям к измерениям: точности и воспроизводимости.

Ключевые слова: метод измерения функционального размера COSMIC; функциональные точки COSMIC (CFP); измерение функционального размера; большая языковая модель (LLM); чат-бот chatGPT; программная инженерия; искусственный интеллект; автоматизация.

Для цитирования: Вальдес-Соуто Ф., Торрес-Робледо Д. Может ли языковая модель на базе chatGPT измерять функциональный размер методом COSMIC? Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 103–114 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–6.

1. Introduction

The competitive software development industry can address the estimating project problem by assessing functional size using a standard (COSMIC ISO/IEC 19761 is the only second-generation FSMM [1]); several estimation methods have been established, such as [2-3], and software development productivity can be measured.

For more than 70 years of research, software estimation has been a focus for numerous researchers since its inception in the 1950's [4]. Precise estimation is a crucial component of software development and a key factor in project failure and has a significant effect on project planning and industrial budgets. [5-6], and [7].

The idea of automating software development chores with artificial intelligence (AI) has gained traction in recent years. Software functional sizing and estimation is one field in which artificial intelligence tries to demonstrate practical and accurate use. [6, 8-9]

The objective is to shorten the time required for measuring using the standard Functional Size Measurement Methods (FSMM). This will enable businesses to quickly estimate functional sizes by having the ability to measure user requirements accurately and promptly, most often provided in text format.

ChatGPT is one of the most advanced models of AI technology, offering some amazing and useful solutions in many fields, such as marketing [10], book creation/edition [11-12], graphic design [13], video creation/edition [14], music edition [15], and so forth. However, not every use has been effective; some attempts have led to pertinent failures or even instances of plagiarism [16].

In this article, we unbiasedly examine whether it is feasible to measure user requirements using ChatGPT 4o and determine whether it is not by providing a specific prompt that outlines the fundamentals of the COSMIC technique.

This paper's outline is as follows. Background information on software estimates and measurement, large language models, functional size measurement, and measurement repeatability is given in Section 2. The experimental protocol and its implementation are explained in Section 3. The data acquired were covered in Section 4, and the conclusions are finally covered in Section 5. Table 1 has the functional size measurement of the user requirements using the COSMIC method.

2. Background

2.1. Software measurement and estimation

The literature on software estimation has a wide range of techniques developed over more than six decades [4]. This has resulted in several estimation methods [2-3, 6-7], numerous classifications of these methods [5-7, 17-19], and various estimation process topologies [20-21]. Despite this extensive catalog of techniques, there is still no consensus on a single model that consistently produces accurate results for all industrial projects.

Even though regression-based estimating techniques based on reference databases predominate in the literature, it is not uncommon to find it difficult to reproduce research [9, 17, 22]. Several authors point out that measuring the size of the program is essential to the precision of approximations [23-26]. According to Fedotova et al. [4], the lack of a size variable may contribute to regression-based models' inability to perform well in estimation.

Neural Networks (NN) and other Machine Learning (ML) techniques have proven to be highly effective in producing accurate predictions, even in situations where noise has severely distorted the input data and the relationships between the inputs and outputs are complex [18].

The academic literature points to several difficulties in the subject, chief among them being the scarcity of real-world datasets (such as those from NASA, ISBSG, Desharnais, and COCOMO) [6]. The use of AI techniques is significantly hampered by this lack.

In the reviewed literature, only two approaches using AI to measure functional size were found, both employing the COSMIC standard [8, 27]. Ungan [8] has presented a technology that measures user requirements based on free-form text. To attain a "precise" measure, it necessitates clear and high-quality specifications, which makes it a closed source. Free-form requirements are by definition prone to being ambiguous, long, and incomplete, especially in the early stages of a project [28].

The other method involves measuring a reference case study using ChatGPT for the first time. This approach yields less than ideal results when applying COSMIC principles and rules; even in cases where the sizes are comparable, there are numerous errors in identifying data groups, data movements, and functional users based on the prompt requirements [27].

2.2 The LLM model: ChatGPT

A Large Language Model (LLM) is an artificial intelligence (AI) model designed to understand and generate human-like texts. LLMs are typically based on deep learning architectures, such as Transformers, and are trained on large amounts of text data to learn the patterns and structures of natural language. These models can perform a variety of language related tasks, including text generation, language translation, question answering, summarization, and more [29].

LLMs have demonstrated remarkable capabilities in "understanding" and generating text across different languages and domains. They are widely used in various applications, such as virtual

assistants, chatbots, content generation, language translation services, and natural language processing tasks Examples of popular LLMs include OpenAI's GPT series (such as GPT-3, GPT3.5, and GPT-4) and Google's BERT.

ChatGPT is an LLM developed by OpenAI, specifically based on the GPT (Generative Pre-trained Transformer) architecture. It uses artificial intelligence to generate responses in text conversations [16]. The functioning of LLMs is based on two main phases: training and fine-tuning.

Text generation in LLMs, like ChatGPT, ***is based on the model's ability to predict the next word in a text sequence*** [16]. When given an input, the model evaluates the previous words and generates a list of possible next words and their associated probabilities. The word with the highest probability is selected, and the process repeats until the response is complete.

Considering the above, LLMs can perform ***a form of reasoning based on statistical and contextual patterns*** learned during training. The models do not have understanding or awareness but operate based on correlations and patterns in the training data.

2.3. Measurement of the functional size of software using COSMIC

Functional Size Measurement Methods (FSMM) are currently divided into two generations [1], with COSMIC ISO/IEC 19761 [28] being the only second-generation FSMM, the lessons learned from first-generation methodologies were the foundation for developing this standard [30]. The COSMIC Measurement Manual [28] presents all the guidelines, precepts, and examples required to carry out functional size measurements.

In real-world projects, approximating a functional size can be necessary in several situations [31]. These include: (1) when a size is required but not enough time or resources are available to measure using the standard method; (2) early in the project's life cycle, before the Functional User Requirements (FUR) have been detailed to the point where an accurate size measurement is possible; and (3) when the documentation quality of the actual requirements is inadequate for an accurate measurement. The functional size assessment can then be as accurate as feasible by using assumptions [32].

2.4. Reproducibility importance in metrology

In any scientific discipline, the validation of results is indispensable. Reproducibility allows other researchers to verify the findings of a study by replicating the same experiment or measurement under the same conditions. If the results can be reproduced, it reinforces the credibility and validity of the original work. This is particularly important in metrology, where the precision and accuracy of measurements directly impact the quality of technological products and services.

The reproducibility thus is crucial for the advancement of the discipline, actually software engineering is considered immature [1, 33].

3. Experimental procedure

To conduct this experiment, we utilized version 4o of ChatGPT (dated 08-08-2024) and the user requirements information from the C-Reg case study [34]. This case study provides the functional size measurement of the requirements for an actual system using the COSMIC method. For reference, the functional size of the C-Reg system is detailed in Table 1.

Fig. 1 illustrates the flow of the experimental procedure to obtain the results presented in Table 2. The first step involved creating a fine-tuned prompt based on the recommendations proposed by [16], which included incorporating some knowledge about COSMIC to be considered by ChatGPT.

The second step is to include the FUR for three functional processes described in the C-Reg system into the prompt and executed it using ChatGPT two times each.

Finally, we compare the measured size using chatGPT against the real measurements obtained in [34].

Table 1. C-Reg case study example for measuring COSMIC function points from requirements [34].

ID	*Functional Process*	CFP
1	Add teacher details	4
2	Consult a teacher's data	4
3	Modify a teacher's data	3
4	Delete data from a teacher	5
5	Consult the Course Offerings (Teacher)	7
6	Create assignments in Course Offerings	6
7	Modify assignments in Course Offerings	7
8	Delete assignments in Course Offerings	4
9	Add student data	4
10	Consult student data	4
11	Modify student data	3
12	Delete student data	4
13	Consult the Course Offerings (Student)	6
14	Create student schedule	6
15	Modify student schedule	8
16	Delete student schedule	6
17	Monitor Course Offering enrollment progress	6
18	Monitor enrollment progress on student schedules	5
19	Close registration	10
	Total	102

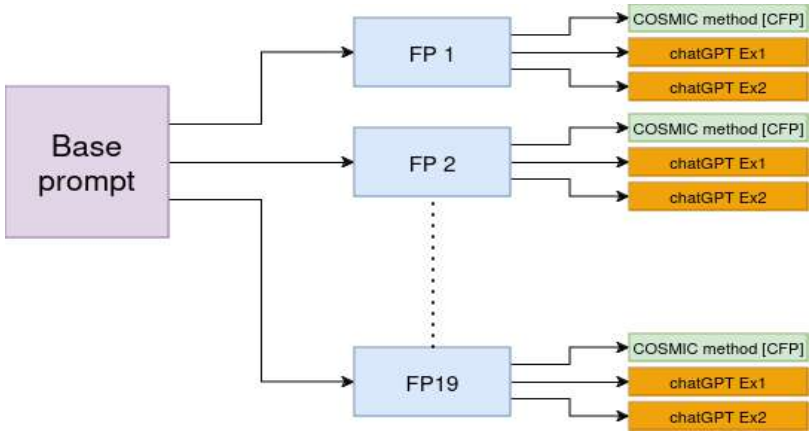


Fig. 1. Flow of experimental procedure diagram.

3.1 Fine-tuned prompt creation

To generate the prompt, it was necessary to describe some aspects of the COSMIC measurement method [28]. Firstly, we defined data groups, data movements, and the definition of functional users.

Creating and fine-tuning the ChatGPT prompt to create an effective functional size measurer involves carefully designing the model to categorize objects of interest, functional users, and the types of data movements [16].

Firstly, we ask to describe the given use case and to include the data groups and their movements according to the COSMIC measurement method [28].

Next, we describe and give examples of what a data movement is, what a functional user is, and what an object of interest is. The examples of these concepts improved how chatGPT classified the content of the use case.

Next, we gave hints on what systems mentioned in the requirements are beyond the measurement scope and must be considered as functional users (Course catalog system and Billing system).

Then there is a space where it is needed to insert any use case.

Finally, we ask that the data groups used, and their movements be explicitly included according to the COSMIC standard, count the times the data groups are moved in the functional process described above, and place them in a table.

Below is the prompt that was used to conduct the tests:

> *Describe the given case use and explicitly include the data groups used and their movements according to the COSMIC standard, the data movements move a group of data and can be read from the database, writing to the database, input from a functional user and output to a functional user, functional users are everything with which the system interacts (e.g. people who use the system, other systems with which it communicates, different systems from which it receives data), data groups describe an object of interest that can be a real world object or a conceptual object (e.g. user, payroll, catalogs, teachers, students, courses, workers).*
>
> *The Course Catalog system and the Billing System are functional users, so to interact with them, there must be exit and entry movements with these.*
>
> *Given the following use case:*
>
> ***[Insert a Functional Process description from C-reg (Table 1)]***
>
> *Explicitly include the data groups used and their movements according to the COSMIC standard, the movements move a group of data and can be read from the database, written to the database, entry from a functional user, and output to a functional user (e. g. people who use the system, other systems with which it communicates), data groups describe an object of interest that can be a real-world object or a conceptual object (e. g. user, payroll).*
>
> *Additionally, count the times the data groups are moved in the functional process described above and place them in a table with the form: data group, movement, value.*

3.2 Prompts execution

Once the execution of distinct prompts was developed, the results were collected and shown in Table 2.

The functional process ID is shown in the first column, and the name of the selected functional process is shown in the second. The functional size derived from the COSMIC technique, as per Table 1, is shown in column three. The functional size obtained from the first prompt execution utilizing the developed prompt and the Magnitude of Relative Error (MRE) for that first prompt execution is displayed in column six.

Table 2. Comparison of the results of applying the COSMIC method to the C-Reg [34] requirements against the results obtained by chatGPT in two different executions.

ID	*Functional Process*	Measured CFP	Measured CFP (ChatGPT)	Diff	MRE
1	Add teacher details (Prompt1)	4	7	3	75.0%
1	Add teacher details (Promp2)	4	10	6	150%
5	Consult the Course Offerings (Teacher) (Prompt1)	7	6	1	14.2%
5	Consult the Course Offerings (Teacher) (Prompt2)	7	10	3	42.8%
15	Modify student schedule (Prompt1)	8	10	2	25.0%
15	Modify student schedule (Promp2)	8	6	2	25.0%

Using the information in Table 2, Table 3 contains the quality criteria for estimation the robustness of the model, which are Mean Magnitude of Relative Error (MMRE), MRE Standard Deviation (SDRMS), and the Prediction level at 10% (Pred 10%) was compiled.

Table 3. Quality criteria for estimation the robustness of the model.

MMRE:	0.553
RMSE:	3.240
SDRMS:	0.510
Pred(10%):	0

Based on the quality criteria, it can be mentioned that there is an average relative error of 5.53%, with a standard deviation of 0.510, and all the measures by ChatGPT are not within the 10% prediction level.

The results show that the size measured using ChatGPT 4o has a difference greater than 10% from the real size measured with the COSMIC method in every FP.

Since the COSMIC method is a standard, the goal of a functional size measure is to be reproducible and audited, any difference could put at risk the project success since there will be a difference in estimating the necessary resources.

4. Discussion

From Table 2, it was observed that like the findings in the article by Hartenstein et al. [27], ChatGPT exhibits some consistency in the total measurement value, in this experiment (only three functional processes) with variation. However, at the individual level, the functional processes yield different

results, while functional process 1 shows a significant percentage variation. This observation suggests that there is no reproducibility based solely on the measurement value.

But after carefully examining ChatGPT's responses, we can see that – even in cases where the definition of the text has been given – it is inconsistent to identify data groups, data movements, and functional users straight from the text – elements that software measurers are familiar with. Refer to Table 4.

Table 4. Two ChatGPT functional size responses for the Add teacher details functional process.

Response 1			*Response 2*		
Data Group	*Movement*	*Value*	*Data Group*	*Movement*	*Value*
Command	*Entry*	*1*	*Command*	*Entry*	*1*
Form Template	*Exit*	*1*	*Form Template*	*Exit*	*1*
Teacher Data	*Entry*	*1*	*Teacher Data*	*Entry*	*1*
Teacher Data	*Read*	*1*	*Teacher Data*	*Read*	*1*
Error Messages	*Exit*	*2*	*Error Messages*	*Exit*	*2*
Teacher Data	*Write*	*1*	*Teacher Data*	*Write*	*1*
			Work Area Data	*Write*	*1*
			Course Catalog Data	*Exit*	*1*
			Billing Data	*Exit*	*1*
Total 7 CFP			*Total 10 CFP*		

One of the primary reasons for using a standard metric is to enable auditability of results. However, in this case, even though the results could be analyzed for accuracy (quality criteria like MMRE, STDEV, etc.), they cannot be audited due to the varying elements used to derive the size. Therefore, it is not possible to consider these results as measurements. At best, they could be considered an approximation approach with some considerations.

As an approximation approach, it has not been studied as extensively as other methods. The results do not provide a clear route for gathering or improvement; they seem more like guesses or luck.

From the observations made in this experiment, it becomes apparent that any approximation approach based on text may encounter similar challenges. It is easy to understand this because LLM models like ChatGPT can perform a form of reasoning based on statistical and contextual patterns learned during training. This implies that there should be identifiable patterns or repeated elements in a text that the model can recognize and utilize. However, this is an open question in the software requirement research field, predating the existence of LLM or natural language processing (NLP) technology.

Additionally, numerous subjective elements such as different local expressions, language variations, communication styles, abstractions, etc., could make this a more difficult task. These factors contribute to the complexity of accurately interpreting and analyzing text-based data, making it challenging to develop robust approximation approaches in software requirements. Therefore, addressing these challenges will require interdisciplinary efforts and innovative solutions to advance the state of the art in software measurement and estimation.

5. Conclusions

In this study, we proposed an experiment to determine whether ChatGPT could perform COSMIC measurements. However, we discovered that ChatGPT could not reliably produce accurate findings at the detailed level. The primary shortcomings found in ChatGPT include its incapacity to accurately extract data movements, data groups, and functional users from the text.

From this experiment, we observe that the results produced by the ChatGPT model are not consistent (reproducibly), leading to different results in different executions. This inconsistency generates

erroneous measurements and incorrect information, a principal element for metrology, which could put at risk the success of the project.

We can observe that ChatGPT does not adhere to the COSMIC methodology, so the resulting measurements, although correct in some cases, are merely coincidental. If the measurement were audited, it would likely not pass the COSMIC method application, which is a significant issue in software contracting.

Based on the experiment, it is challenging to conceive that LLM models could accurately measure software using a FSMM because they operate on patterns and structured data. In contrast, FURs are often described in free text and depend on the individual writer, leading to variations in language, communication styles, and abstractions.

Indeed, our challenge extends beyond the capabilities of current technology like ChatGPT, it includes the inherent variability and subjectivity of human language. Despite the significant advancements in natural language processing, there is no replicable or consistent way to interpret Functional Unit Requirements (FURs) due to the lack of standardized descriptions.

The proposal by Gérançon et al. [35] offers a potential approach to address this issue. However, current tools, even the most advanced ones like ChatGPT, fall short in meeting the essential requirements for measurement using the COSMIC method directly from the FURs text descriptions: accuracy and reproducibility.

Therefore, it's crucial to recognize that the limitations and issues faced in using AI for estimation tasks extend to approximating functional size from text. Addressing these challenges requires advancements in AI technology, as well as a deeper understanding of the complexities of human language and the specific requirements of the software engineering domain.

6. Limitations

Since OpenAI's ChatGPT model, is not open source, and they can change the way the model responds to mitigate risky results according to their policy [29], the quality of the results in this proposal has been changing from the beginning of this article until the time of its publication (a reproducibility problem).

The steps to perform a functional size measurement using the COSMIC method require knowing the context of the system to be measured, such as the attributes of an object of interest. Thus, the way an LLM will group the attributes of an object of interest could be different from those identified in the measurement process.

How user requirements are obtained can vary, so it is a challenge to create a prompt that could cover everything.

7. Future work

Replicating the experiment with all the functional processes from the case study [34], along with additional real software applications, would provide valuable insights into the performance and limitations of ChatGPT in measuring functional size from text.

Using different or improved versions of LLM, such as ChatGPT 4, 4+, 4o, LLAMA 3.1, Gemini, and Bart, could offer further insights into the model's consistency and performance. To get consistent results under the same prompt in any model would be crucial in determining its reliability.

Finally, exploring the development of a system that leverages ChatGPT for specific parts of the measurement methodology, such as transforming user stories into functional user actions or wrapping attributes in objects of interest, could lead to creating a more accurate measurement model. A mixed approach, like the Retrieval-Augmented Generation (RAG) approach, could combine the strengths of ChatGPT's language processing capabilities with other techniques or models to enhance accuracy and reliability in functional size measurement.

References

[1]. ABRAN, Alain. Software metrics and software metrology. John Wiley & Sons, 2010. https://doi.org/10.1002/9780470606834.ch2.

[2]. Silhavy, R., Prokopova, Z. & Silhavy, P. Algorithmic optimization method for effort estimation. Program Comput Soft 42, 161–166 (2016). https://doi.org/10.1134/S0361768816030087.

[3]. Durán, M., Juárez-Ramírez, R., Jiménez, S. et al. User Story Estimation Based on the Complexity Decomposition Using Bayesian Networks. Program Comput Soft 46, 569–583 (2020). https://doi.org/10.1134/S0361768820080095.

[4]. O. Fedotova, L. Teixeira, A.H. Alvelos, Software effort estimation with multiple linear regression: Review and practical application, J. Inf. Sci. Eng. 29 (2013) 925–945.

[5]. T.K. Lee, K.T. Wei, A.A.A. Ghani, Systematic literature review on effort estimation for Open Sources (OSS) web application development, in: FTC 2016 - Proc. Futur. Technol. Conf., IEEE, San Francisco, California, USA, 2016: pp. 1158–1167. https://doi.org/10.1109/FTC.2016.7821748.

[6]. P. Sharma, J. Singh, Systematic literature review on software effort estimation using machine learning approaches, in: Proc. - 2017 Int. Conf. Next Gener. Comput. Inf. Syst. ICNGCIS 2017, IEEE, Jammu, India, 2018: pp. 54–57. https://doi.org/10.1109/ICNGCIS.2017.33.

[7]. C.E. Carbonera, K. Farias, V. Bischoff, Software development effort estimation: A systematic mapping study, IET Res. Journals. 14 (2020) 1–14. https://doi.org/10.1049/iet-sen.2018.5334.

[8]. E. Ungan, C. Hammond, A. Abran, Automated COSMIC Measurement and Requirement Quality Improvement Through ScopeMaster ® Tool, in: A.C. Murat Salmanoglu (Ed.), Proc. Acad. Pap. IWSM Mensura 2018 "COSMIC Funct. Points - Fundam. Softw. Effort Estim. Held Conjunction with China Softw. Cost Meas. Conf. (CSCM 2018), CEUR Workshop Proceedings (CEURWS.org), Beijing, China, 2018: pp. 1–13. doi: ISSN:1613-0073.

[9]. P. L. Braga, A. L. I. Oliveira and S. R. L. Meira, "Software Effort Estimation using Machine Learning Techniques with Robust Confidence Intervals," in 7th International Conference on Hybrid Intelligent Systems, Kaiserslautern, Germany, 2007.

[10]. Yaozhi Zhang, Nina Katrine Prebensen, Co-creating with ChatGPT for tourism marketing materials, Annals of Tourism Research Empirical Insights, Volume 5, Issue 1, 2024, 100124, ISSN 2666-9579, https://doi.org/10.1016/j.annale.2024.100124.

[11]. Altmäe, Signe Sola-Leyva, Alberto Salumets, Andres, Artificial intelligence in scientific writing: a friend or a foe?, Volume 47, Issue 1, 2023, ISSN 1472-6483 https://doi.org/10.1016/j.rbmo.2023.04.009.

[12]. Zuckerman, M., Flood, R., Tan, R. J. B., Kelp, N., Ecker, D. J., Menke, J., & Lockspeiser, T. (2023). ChatGPT for assessment writing. Medical Teacher, 45(11), 1224–1227. https://doi.org/10.1080/0142159X.2023.2249239.

[13]. T. Putjorn and P. Putjorn, "Augmented Imagination: Exploring Generative AI from the Perspectives of Young Learners," 2023 15th International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thailand, 2023, pp. 353-358, doi: 10.1109/ICITEE59582.2023.10317680.

[14]. S. Bengesi et al., Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers. arXiv preprint arXiv:2311.10242 (2023).

[15]. McKinsey & Company, What is ChatGPT, DALL-E, and generative AI? | McKinsey. McKinsey & Company (2023).

[16]. OpenAI and Josh Achiam and Steven Adler and Sandhini Agarwal, GPT-4 Technical Report, 2024. arXiv:2303.08774.

[17]. M. Jørgensen, M. Shepperd, A systematic review of software development cost estimation studies, IEEE Trans. Softw. Eng. 33 (2007) 33–53. https://doi.org/10.1109/TSE.2007.256943.

[18]. S. Bilgaiyan, S. Sagnika, S. Mishra, M. Das, A systematic review on software cost estimation in Agile Software Development, J. Eng. Sci. Technol. Rev. 10 (2017) 51–64. https://doi.org/10.25103/jestr.104.08.

[19]. N. Kinoshita, A. Monden, M. Tshunoda and Z. Yucel, "Predictability classification for software effort estimation," in Proceedings - 2018 IEEE/ACIS 3rd International Conference on Big Data, Cloud Computing, Data Science and Engineering, BCD 2018, Yonago, Japan, 2018.

[20]. R. Britto, V. Freitas, E. Mendes, M. Usman, Effort estimation in global software development: A systematic literature review, Proc. - 2014 IEEE 9th Int. Conf. Glob. Softw. Eng. ICGSE 2014. (2014) 135–144. https://doi.org/10.1109/ICGSE.2014.11.

[21]. F. Valdés-Souto, Validation of supplier estimates using cosmic method, CEURInternational Work. Softw. Meas. Int. Conf. Softw. Process Prod. Meas. (IWSM Mensura 2019). 2476 (2019) 15–30.

[22]. M. Shin and A. L. Goel, "Empirical Data Modeling in Software Engineering Using Radial Basis Functions," IEEE Transactions on Software Engineering, vol. 26, no. 6, pp. 567-576, 2000.
[23]. M. Linda and M. C. B. Laird, Software Measurement and Estimation: A Practical Approach, New York, N.Y., USA: Jonh Wiley & Sons, 2006.
[24]. S. Koch and J. Mitlöhner, "Software project effort estimation with voting rules," Decision Support Systems, vol. 46, no. 4, pp. 895-901, 2009.
[25]. De Lucia, E. Pompella and S. Stefanucci, "Assessing effort estimation models for corrective maintenance through empirical studies," Information and Software Technology, vol. 47, no. 1, pp. 3-15, 2005.
[26]. J. Hill, L. C. Thomas and D. E. Allen, «Experts' estimates of task durations in software development projects», International Journal of Project Management, vol. 18, nº 1, pp. 13-21, 2000.
[27]. Hartenstein, S., Johnson, S.L., Schmietendorf, A., ¨Towards a fast cost estimation Supported by large language models¨ (2024). URL: https://cosmic-sizing.org/publications/fast-cost-estimation-by-chatgpt/
[28]. The COSMIC Functional Size Measurement Method: Measurement Manual (2021), v. 5.0 ed., URL https://cosmic-sizing.org/measurement-manual/
[29]. OpenAI and Josh Achiam and Steven Adler and Sandhini Agarwal, GPT-4 System Card, 2024. arXiv:2303.08774
[30]. F. Vogelezang and H. v. Heeringen, Benchmarking: Comparing Apples to Apples (Apress, Berkeley, CA, 2019), pp. 205–217, ISBN 978-1-4842-4221-6.
[31]. Vogelezang, COSMIC Group, ¨Early Software Sizing with COSMIC, Practitioners¨ (2020), v.4.0.2, URL: https://cosmic-sizing.org/publications/early-software-sizing-with-cosmic-practitioners-guide/
[32]. Vogelezang, COSMIC Group, ¨Early Software Sizing with COSMIC: Experts Guide¨ (2020), v.4.0.2, URL: https://cosmic-sizing.org/publications/early-software-sizing-with-cosmic-experts-guide/
[33]. Sánchez Alonso, S., Sicilia Urban, M. Á., & Rodríguez García, D. (2011). Ingeniería del software : un enfoque desde la guía SWEBOK (1a ed., 1a reimp.). Garceta.
[34]. Symons, C.R., et al, Course Registration ('C-REG') System Case Study, v2.0.1 2018. https://cosmic-sizing.org/publications/course-registration-c-reg-system-case-study-v2-0-1/
[35]. Bruel Gérançon, Sylvie Trudel, Roger Kkambou, Serge Robert, Software Functional Sizing Automation from Requirements Written as Triplets, ICSEA 2021: The Sixteenth International Conference on Software Engineering Advances, 2021.

Информация об авторах / Information about authors

Франсиско ВАЛЬДЕС-СОУТО имеет степень PhD в области инженерии программного обеспечения по специальности "Измерение и оценка программного обеспечения" в Высшей технологической школе (ETS) в Канаде, две магистерские степени в Мексике и Франции. Президент COSMIC. Доцент факультета наук Национального автономного университета Мексики. Основатель Мексиканской ассоциации метрик программного обеспечения (AMMS). Более 25 лет опыта в разработке критически важного программного обеспечения. К настоящему времени опубликовал более 50 научных работ, включая статьи в индексированных журналах, трудах научных конференций, книгах и главах книг. Является главным промоутером проекта изучения формальных метрик программного обеспечения в Мексике, продвигая COSMIC (ISO/IEC 19761) в качестве национального стандарта. Член Национальной системы исследователей (SNI). Область научных интересов: измерение и оценка программного обеспечения, применяемого для управления проектами программного обеспечения, управление тематикой, производительностью и экономикой разработок программного обеспечения.

Francisco VALDÉS-SOUTO had a PhD in Software Engineering with a specialty in Software Measurement and Estimation at the École de Technologie Supérieure (ETS) in Canada, two master's degrees in Mexico and France. President of COSMIC. Associate Professor of the Faculty of Sciences of the National Autonomous University of Mexico (UNAM). Founder of the Mexican Association of Software Metrics (AMMS). More than 25 years of experience in critical software development. He currently has more than 50 publications including articles in Indexed Journals, Proceedings, books and book chapters. He is the main promoter of the topic of formal software metrics in Mexico, promoting COSMIC (ISO/IEC 19761) as a National Standard. Member of the National System of

Researchers (SNI). Research interests: software measurement and estimation applied to software project management, scope management, productivity and economics in software projects.

Даниэль ТОРРЕС-РОБЛЕДО – магистрант Исследовательского института в области прикладной математики и систем, имеет ученую степень по программированию от научного факультета Национального автономного университета Мексики.

Daniel TORRES-ROBLEDO – Master student at Research Institute in Applied Mathematics and Systems, degree in Computer Science from Science Faculty of the UNAM.

DOI: 10.15514/ISPRAS-2024-36(6)-7

Технические и социальные навыки, необходимые для квантовых вычислений

[1] *Р. Хуарес-Рамирес, ORCID: 0000-0002-5825-2433 <reyesjua@uabc.edu.mx>*
[1] *С. Хименес, ORCID: 0000-0003-0938-7291 <samantha.jimenez@uabc.edu.mx>*
[1] *К. Х. Наварро, ORCID: 0000-0002-7220-7006 <cnavarro@uabc.edu.mx>*
[2] *С. Гуерра-Гарсия, ORCID: 0000-0002-9290-6170 <cesar.guerra@uaslp.mx>*
[2] *Э. Г. Перес-Гонзалес, ORCID: 0000-0003-3331-2230 <hectorgerardo@uaslp.mx>*
[3] *К. Фернандес-и-Фернандес, ORCID: 0000-0002-1586-8772 <caff@mixteco.utm.mx>*
[4] *Х. Ортис-Эрнандес, ORCID: 0000-0002-6481-1537 <javier.oh@cenidet.tecnm.mx>*
[5] *К. Кансино, ORCID: 0000-0002-7467-9082 <karina.cancino@uptapachula.edu.mx>*

[1] *Автономный университет Нижней-Калифорнии,*
Нижняя Калифорния, Мексика.
[2] *Автономный университет Сан Луис Потоси,*
Сан Луис Потоси, Мексика.
[3] *Технологический университет Миштека,*
Оахака, Мексика.
[4] *Национальный центр исследований и технологических разработок CENIDET,*
Куэрнавака, Мексика.
[5] *Технологический университет Тапачула,*
Тапачула, Мексика.

Аннотация. Квантовые вычисления, существенным образом основанные на квантовой механике, представляют значительные проблемы для людей, не имеющих достаточных знаний квантовой физики. Для работы в этой области необходимы значительные научные познания в квантовой теории, математике и смежных областях. Кроме того, междисциплинарный характер квантовых вычислений для эффективной командной работы требует наличия различных социальных навыков. В этой статье рассматривается и систематизируется научная литература, выявляя ключевые технические и социальные навыки, необходимые для подготовки студентов и специалистов в области квантовых вычислений, и помогая учебным заведениям разрабатывать соответствующие курсы и учебные планы.

Ключевые слова: квантовые вычисления; квантовая механика; математика; научные знания; профессиональные программистские навыки; учебное планирование.

Для цитирования: Хуарес-Рамирес Р., Хименес С., Наварро К. Х., Гуерра-Гарсия С., Перес-Гонзалес Э. Г., Фернандес-и-Фернандес К., Ортис-Эрнандес Х., Кансино К. Технические и социальные навыки, необходимые для квантовых вычислений. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 115–148. DOI: 10.15514/ISPRAS–2024–36(6)–7.

Благодарности. Авторы благодарят Математический исследовательский центр (CIMAT A.C.) за поддержку в предоставлении доступа к электронным базам данных. Кроме того, авторы благодарны Автономному университету Нижней Калифорнии за финансирование этого исследования в рамках проекта 3908_300/6/C/63/23 «Computación Cuántica: Implicaciones en la Ingeniería de Software y las Competencias del Ingeniero de Software», 23a Convocatoria Interna de Proyectos de Investigation ación. Наконец, авторы благодарят команду Мексиканского общества программной инженерии (REDMIS), которая участвовала в этом проекте и обзоре литературы.

Technical and Soft Skills Required for Quantum Computing

[1] *R. Juárez-Ramírez, ORCID: 0000-0002-5825-2433 <reyesjua@uabc.edu.mx>*
[1] *S. Jiménez, ORCID: 0000-0003-0938-7291 <samantha.jimenez@uabc.edu.mx>*
[1] *C. X. Navarro, ORCID: 0000-0002-7220-7006 <cnavarro@uabc.edu.mx>*
[2] *C. Guerra-García, ORCID: 0000-0002-9290-6170 <cesar.guerra@uaslp.mx>*
[2] *H. G. Perez-Gonzalez, ORCID: 0000-0003-3331-2230 <hectorgerardo@uaslp.mx>*
[3] *C. Fernández-y-Fernández, ORCID: 0000-0002-1586-8772 <caff@mixteco.utm.mx>*
[4] *J. Ortiz-Hernández, ORCID: 0000-0002-6481-1537 <javier.oh@cenidet.tecnm.mx>*
[5] *K. Cancino, ORCID: 0000-0002-7467-9082 <karina.cancino@uptapachula.edu.mx>*

[1] *Universidad Autónoma de Baja California,*
Baja California, México.
[2] *Universidad Autónoma de San Luis Potosí,*
San Luis Potosí, México.
[3] *Universidad Tecnológica de la Mixteca,*
Oaxaca, México.
[4] *Centro Nacional de Investigación y Desarrollo Tecnológico CENIDET,*
Cuernavaca, México.
[5] *Universidad Tecnológica de Tapachula,*
Tapachula, México.

Abstract. Quantum computing, strongly based on quantum mechanics, presents significant challenges for people without a background in quantum physics. Strong technical skills in quantum topics, mathematics, and related fields are essential for job roles in this area. Additionally, the multidisciplinary nature of quantum computing requires soft skills for effective teamwork. This paper reviews literature using systematic mapping to identify the key technical and soft skills needed to prepare students and professionals for the quantum computing field, helping educational institutions design appropriate courses and curricula.

Keywords: quantum computing; quantum mechanics; mathematics; technical skills; soft skills; curricula.

For citation: Juárez-Ramírez R., Jiménez S., Navarro C. X., Guerra-García C., Perez-Gonzalez H. G., Fernández-y-Fernández C., Ortiz-Hernández J., Cancino K. Technical and soft skills required for quantum computing. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 115-148 (in Russian). DOI: 10.15514/ISPRAS-2024-36(6)-7.

Acknowledgements. We would like to thank Centro de Investigación en Matemáticas (CIMAT A.C.) for their support in providing access to electronic databases. Also, we would like to thank Universidad Autónoma de Baja California, for financing this research through the Project 3908_300/6/C/63/23 “Computación Cuántica: Implicaciones en la Ingeniería de Software y las Competencias del Ingeniero de Software”, 23a Convocatoria Interna de Proyectos de Investigación. Finally, thanks to the team Red Mexicana de Ingeniería de Software (REDMIS) who participated in this project and literature review.

1. Введение

Квантовые вычисления (QC) в настоящее время являются одним из самых многообещающих достижений, которые должны радикально изменить вычислительные машины и вычисления на них [1-9]. Построенные на принципах квантовой механики, таких как суперпозиция, запутанность и измерение, квантовые вычисления могут открыть перед нами совершенно новые перспективы. Квантовые вычисления неотделимы от квантовой механики [9-11]; их реализация требует и математического фундамента, и понимания вероятностного характера кубитов, и того, как можно ими управлять. Для тех, кто специализируются в области теории и практики вычислений на вычислительных машинах, а особенно для инженеров-программистов переход к использованию квантовых компьютеров представляет серьезные трудности.

Квантовая природа квантовых вычислений делает их недоступными для понимания специалистами, не работающими в области физики. Квантовые вычисления испытывают мощное влияние как квантовой механики, так и математики. Они включают в себя ключевые понятия из этих наук, в том числе следующие [5, 10, 12-13]:

1) квантовая механика – суперпозиция, запутанность, квантовая интерференция, измерение;
2) квантовые кубиты и квантовые состояния;
3) квантовые вентили и квантовые схемы;
4) линейная алгебра (векторные пространства, матрицы, собственные значения и собственные векторы);
5) комплексные числа;
6) теория вероятностей;
7) тензорное исчисление;
8) преобразования Фурье.

Математический базис квантовой механики, изучающей поведение элементарных частиц материи, был первоначально создан в начале XX века Нильсом Бором [14], Вернером Гейзенбергом [15], Эрвином Шредингером [16], Полем Дираком [17] и другими. В настоящее время в основе квантовых вычислений находятся именно эти фундаментальные принципы.

Вся необходимая совокупность знаний делает квантовые вычисления более сложными, чем классические. В результате программисты могут столкнуться с проблемами при попытке осмысливать и развивать квантовое программное обеспечение.

Многодисциплинарный подход к квантовым вычислениям. Смежными с квантовыми вычислениями областями являются квантовая информатика (QIS) и квантовые информационные технологии (QIST). Эти две научные области сильно взаимосвязаны, причем последняя делает основной упор на технологию разработки и управления. Сочетание в квантовой информатике элементов теории информации, информатики и квантовой механики способно привести к прорыву в области обработки информации [18].

Внедрением новых квантовых технологий в таких областях, как связь, сети, безопасность данных, навигация и медицинская диагностика [11, 19-20] с использованием достижений квантовой информатики заняты междисциплинарные коллективы. Квантовая информатика, сочетающая в себе квантовые принципы, формирует у разработчиков программного обеспечения новое мышление.

Текущий дефицит специалистов, прошедших обучение квантовым вычислениям. В последние годы квантовые вычисления становятся все более и более важными для всех наук, связанных с вычислительными машинами, и смежных с ними (например, с прикладной физикой [21] или вычислительной химией [22]). Развитию новых технологий и методов посвящено значительное количество ведущихся исследований [23]. Значительную трудность представляет высокая сложность квантовых вычислений, особенно в том смысле, который придают этому слову математика и физика, а это делает такие вычисления менее доступными для тех, кто работает в области теории и практики программирования [24].

Нехватка квалифицированных специалистов в области квантовых вычислений совершенно очевидна. В инфраструктурных скриптах и связующем коде в настоящее время встречается множество ошибок [25]. Для снижения их числа необходима более заметная языковая поддержка, для разработки квантового программного обеспечения и платформ, совершенно необходимы квалифицированные программисты. В настоящее время разработчиками квантовых программ становятся выпускники ВУЗов, не имеющие программистского образования, в частности, физики [25-26], которым не хватает серьезного классического образования в программной инженерии, в том числе навыков эффективного интеграционного тестирования.

Исследования показывают наличие пробелов в знаниях квантового программного обеспечения на предприятиях [27], что свидетельствует о необходимости изменений в методах программной инженерии и образовании [23].

Пути решения проблемы. Разработка квантовых программ – это динамично растущая отрасль, ориентированная на использование возможностей квантовых вычислений для различных приложений [3, 28-32]. В квантовых компьютерах используются кубиты, которые могут существовать в суперпозициях двух состояний, что по сравнению с классическими компьютерами позволяет быстрее и эффективнее выполнять сложные вычисления [30]. Однако разработка квантового программного обеспечения сталкивается с такими проблемами, как шум, необходимость исправления ошибок, масштабируемость и способность к взаимодействию [28-29, 32].

Для решения проблемы нехватки квалифицированных специалистов в отрасли поставщики квантовых технологий должны согласовать свои усилия с принципами квантовой программной инженерии, предлагая достаточный набор программных инструментов и обучающих курсов [3].

Чтобы добиться успеха в своей работе, создатели квантового программного обеспечения должны знать [33]:

1) все детали проблемы, которую они пытаются решить;
2) как написать квантовый код;
3) как записать решение своей проблемы в виде квантовой программы;
4) как писать масштабируемые квантовые решения (в зависимости от сценария использования).

Осознавая все эти требования отрасли, крайне важно определить знания и навыки, необходимые практикующим специалистам по квантовым вычислениям, особенно тем, кто специализируется в области квантовой программной инженерии. Таким образом, это исследование направлено на:

1) определение компетенций, необходимых в индустрии квантового программного обеспечения, с особым вниманием к знаниям и навыкам, существенным для специалистов в области квантовых вычислений.
2) определение ключевых концепций для разработки будущих учебных программ и образовательных мероприятий, способствующих привлечению учащихся к изучению научных основ квантовой информатики.

Важно отметить, что настоящая статья представляет собой несколько сокращенный вариант развернутой статьи “Skills required for Quantum Computing: A comprehensive review of recent Studies”, которая публикуется в этом году, и в которой можно найти более детальную информацию по затронутым вопросам [34].

Оставшаяся часть статьи организована следующим образом. Раздел 2 содержит описание текущей ситуации, включая описание потребностей предприятий-разработчиков в терминах компетенций, навыков и возможностей квантовых технологий. В разделе 3 описываются некоторые близкие работы, в которых представлены результаты подготовки студентов и проверки их компетенций в этой области знаний. В разделе 4 описывается методология, использованная при анализе литературных источников, суммируются шаги, выполнявшиеся для систематизации полученных данных. Раздел 5 представляет результаты выполненного обзора литературы, выделяя три основных элемента: фундаментальные знания для квантовых вычислений, необходимые технические (или профессиональные) навыки и необходимые социальные навыки. В разделе 6 приводится краткое обсуждение результатов обзора литературы и основных идей авторов. В разделе 7 содержатся выводы и будущая работа. В конце статьи представлены раздел ссылок и краткая информация о ее авторах.

2. Предыстория

Учебные заведения должны решать проблему нехватки квалифицированных специалистов в области квантовых вычислений, разрабатывая учебные программы, которые четко определяют компетенции и навыки, необходимые для будущих инженеров-разработчиков квантового программного обеспечения [3]. Эти новые учебные планы должны включать конкретные навыки, необходимые для разработки квантовых программ [35]. Кроме того, для эффективного решения этой проблемы и использования достижений квантовых вычислений необходимо наладить сотрудничество между промышленностью, академическими кругами и заинтересованными государственными структурами [1].

Для решения этой проблемы уже были предприняты определенные усилия. Далее описаны некоторые из них.

Потребности квантовой промышленности. В работе [36] приводится обзор 57 компаний, занятых в квантовой промышленности. Целью работы было определение номенклатуры рабочих мест, навыков и уровня подготовки для нанимаемых работников. Исследование проводилось в помощь образовательным учреждениям, которым было нужно адаптировать свои программы к требованиям отрасли на основе понимания конкретных рабочих ролей, навыков и уровней квалификации, необходимых в квантовом секторе.

Опрос выявил разнообразие возможностей по трудоустройству в квантовой индустрии, от узкоспециализированных ролей, таких как разработчики квантовых алгоритмов, до более широких позиций в бизнесе и программном обеспечении. Компании ищут и обладателей дипломов бакалавров, и кандидатов наук, но стараются нанимать защитивших магистерские и кандидатские диссертации. Выделены семь должностей, которые явно требуют навыков в квантовых вычислениях:

1) специалист по исправлению ошибок;
2) физик-экспериментатор;
3) физик-теоретик;
4) вычислительный химик;
5) инженер/научный сотрудник по фотонике/оптике;
6) разработчик квантовых алгоритмов;
7) архитектор приложений/решений.

Специальности и роли в экосистемах квантовых вычислений. Проанализировав 15 ведущих компаний, ведущих работы над глобальной экосистемой квантовых вычислений (global Quantum Computing Ecosystem), среди которых были компании IBM, Amazon, Microsoft, и Google, автор работы [8] определил, что в целом все изученные роли могут быть распределены по пяти непересекающимся кластерам специальностей:

1) программная инженерия,
2) аппаратная инженерия,
3) научные исследования,
4) технические консультации, и
5) управление продуктами и программами.

В качестве примеров должностей для Кластера 1 можно привести такие:

- инженер-разработчик квантовых программ;
- инженер-разработчик программ для встроенных систем;
- инженер-тестировщик и измеритель квантовых программ.

Работники, относящиеся к этому кластеру, выполняют, например, такие работы:

- развитие технологий квантовых вычислений для полного охвата стека архитектур;

- развитие приложений, чтобы сделать квантовый подход широко используемым и доступным.

Примеры навыков, необходимых для этой категории, включают такие языки, библиотеки, системы и технологии, как C, C++, Java, Python, Spring, MATLAB, HTML, Simulink, PostgreSQL, Julia, создание сценариев использования программного обеспечения для аппаратных интерфейсов.

Отчет, представленный в работе [8], показал, что более универсальные рабочие места в квантовой индустрии требуют сразу многих различных навыков, отнюдь не все из которых напрямую связаны с квантовыми вычислениями. Так как в должностных инструкциях могут отсутствовать точные детали требований, предъявляемых к работникам, для определения конкретных навыков, необходимых для квантовых должностей, предпринимаются и другие усилия, которые обсуждаются в следующем разделе.

3. Обзор аналогичных работ

Быть квантово-готовым означает обладать и теоретическими знаниями, и практическими навыками, нужными для работы с квантовыми технологиями [37]. Образовательные и научные учреждения могут помочь удовлетворить эту потребность, предлагая практикантам специализированные программы, проводя исследования и сотрудничая с промышленностью.

Сложность квантовой механики делает квантовую информатику и квантовые технологии очень трудными для изучения и овладения. Исследования эффективности обучающих курсов QIST ограничены, а отсутствие стандартизированной учебной программы является серьезной проблемой из-за междисциплинарного характера области [38]. Растет признание необходимости учебных программ уже на уровне бакалавриата, как и потребности в учебных материалах по квантовой информатике и квантовой инженерии.

3.1 Усилия в области образования по специальностям квантовых вычислений

В работе [38] для более глубокого понимания поставленных проблем предлагается сочетать вводные лекции по квантовым вычислениям с практическими занятиями. Опрос, проведенный компанией IBM в 2020 году, предлагает включить в программу такие традиционные дисциплины, как линейная алгебра и квантовая механика, а также практические материалы, например, работу с языком Python, платформой Qiskit, а также практикум по калибровке оборудования с обучением студентов методам интеграции квантовых систем. Именно на интеграции реальных квантовых систем в учебную программу сделан упор в рекомендациях компании IBM. Некоторое представление о курсе, предложенном в работе [38], дается в табл. 1. Хотя в исследовании сообщалось о положительных результатах, оно представляет собой всего лишь одиночную инициативу, а не широко воспроизводимую модель.

В работе [13] ее авторы подчеркнули отсутствие устоявшихся традиций и исследований по обучению квантовой физике студентов, не имеющих физического образования. Они предлагают ввести в магистратуре курс, охватывающий восемь ключевых тем по квантовой физике и квантовой вычислительной технике:

1) волновые функции,
2) динамика функций движущихся волн;
3) квантование (не зависящее от времени уравнение Шредингера);
4) квантовая запутанность;
5) специфические квантовые алгоритмы;
6) универсальные вентили и их аппроксимации;

7) квантовые решения задач оптимизации;
8) квантовые вычисления в условиях средней зашумленности.

Табл. 1. Тематика лекционных курсов по квантовым вычислениям, поддержанных практическими занятиями.
Table 1. Quantum computation course topics with hands-on supplementary materials.

Традиционная лекционные курсы	**Практические занятия**
Модуль 1:	
Линейная алгебра	Язык программирования Phyton
Квантовая механикаs	Платформа Qiskit
Нотация Дирака	
***Модуль 2*:**	
Квантовые алгоритмы	Квантовое программирование
***Модуль 3*:**	
Аппаратура квантовых компьютеров	Калибровка квантового оборудования
	Устранение ошибок
	Преобразование программ

Групповые интервью и анкеты показывают, что студенты, изучающие информационные технологии, вполне способны к изучению квантовой физики и заинтересованы в ее изучении в рамках своего образования в области квантовых вычислений. Комплексный подход, сочетающий квантовую физику и информатику в одном курсе, в целом работает хорошо.

В работе [39] подчеркнуто, что квантовые информационные технологии (QIST) по своей сути междисциплинарны, они включают в себя такие области, как физика, информатика и инженерия. В своей работе авторы предложили три учебных плана по интеграции в программы бакалавриата по информатике тематик QIST. Эти планы были разработаны при участии промышленности, научных кругов и правительства. Каждый план включает в себя ряд тем, подходящих для семестровых курсов, но потенциально распространяемых на два семестра с дополнительными лабораторными занятиями. Темы охватывают фундаментальные принципы, кубиты, квантовые вентили, алгоритмы и практические аспекты, такие как квантовый интернет и устранение ошибок.

3.2 Усилия в области образования по специальностям квантовой программной инженерии

В некоторых исследованиях изучалось влияние квантовых вычислений на разработку программного обеспечения, анализировалось, как они влияют на каждую фазу жизненного цикла разработки программного обеспечения и прогресс в адаптации этих фаз к квантовым вычислениям.

Авторы работы [23] определили технические компетенции, необходимые студентам магистратуры по информатике для квантовых вычислений, сравнив их с текущими компетенциями выпускников бакалавриата. Их результаты подчеркивают разрыв между имеющимися и необходимыми компетенциями, предлагая, чтобы такие темы, как математика, физика и теория сложности, были включены в учебные планы по программной инженерии, как например, в стандарте SWEBOK [40], который определяет все этапы разработки программного обеспечения. На основе выполненного анализа компетенций авторы предлагают разработать концепцию обучения, чтобы лучше готовить студентов к работе по созданию квантового программного обеспечения.

В работе [41] авторы делятся своим опытом преподавания курса “Квантовые вычисления” для студентов технических факультетов, начиная со второго по пятый год обучения. Курс

предназначался для студентов, уже имевших знания в области дифференциального исчисления, линейной алгебры и программирования, но не изучавших квантовую механику и современную физику. Курс был направлен на развитие практических навыков квантового программирования, отталкиваясь не от описания базовых физических явлений, а от абстрактной логики квантовых вычислений. Курс использовал язык Q# и был основан на учебной программе, предоставленной квантовой сетью компании Майкрософт (Microsoft Quantum Network). Основными были признаны следующие темы:

1) введение в квантовые вычисления;
2) простые алгоритмы;
3) алгоритм квантового преобразования Фурье (QFT), блочные шифры Simon, оценка фазы и квантовый алгоритм факторизации Шора;
4) квантовый алгоритм решения задачи перебора (алгоритм Гровера);
5) реальность квантовых компьютеров.

3.3 Совершенствование деятельности по разработке программного обеспечения с помощью квантовых алгоритмов

В работе [42] мотивация была подчеркнута прогрессом квантовых компьютеров (КК), которые могут вскоре решить ранее казавшиеся неразрешимыми проблемы в таких областях, как химия, физика и финансы. Однако потенциал квантовых алгоритмов для ускорения задач программной инженерии (SE) остается неизученным. Авторы указанной работы провели исследование восьми групп квантовых алгоритмов, которые потенциально могут ускорить различные задачи программной инженерии на разных этапах, выявляя как новые возможности, так и возникающие проблемы:

1) решатели систем разреженных линейных уравнений;
2) алгоритмы для решения задач нахождения собственных значений и собственных векторов;
3) решатели систем дифференциальных уравнений [43-44];
4) методы подгонки моделей к данным и анализа точности подгонки;
5) квантовое машинное обучение (QML) [45-46];
6) комбинаторная (дискретная) оптимизация [47-49];
7) поиск и сравнение строк [50-53];
8) алгоритмы для проверки выполнимости формул пропозициональной логики в конъюнктивной нормальной форме.

3.4 Классификация работников, занятых квантовыми вычислениями

В работе [41] авторы исследуют классификацию работников, работающих в области квантовых вычислений, распределяя их по шести категориям, предложенным в подкасте [54] (см. табл. 2).

Эта классификация поддержана в работе [55], которая содержит сведения о проведенном опросе, в котором для различных учебных групп определены специфические образовательные подходы. В эти группы входили специалисты, повышающие свою квалификацию, интересующиеся всем новым старшеклассники, соискатели формальной магистерской степень магистра в этой области науки.

Классификация в табл. 2 помогает нам:

1) визуализировать различные уровни квалификации ведущих квантовые вычисления работников и соответственно ранжировать этих работников;

2) разработать курсы по квантовым вычислениям и программы практических занятий, адаптированные к этим уровням, потенциально интегрируя и другие навыки, такие как показанные в работе [56] и традиционной таксономии Блума [57-58].

Табл. 2. Классификация занятых в квантовых вычислениях.
Table 2. Classification of quantum workforce.

Категория	Описание
Квантовое любопытство (Quantum Curious)	Ищущие знаний об основах квантовых вычислений
Квантовое знакомство (Quantum Explorer)	Начавшие изучать некоторые концепции квантовых вычислений
Квантовый рост (Quantum Climber)	Решившие получить образование в области квантовых вычислений
Квантовое подключение (Quantum Enabled)	Знакомые с синтаксисом языков квантового программирования и с программными платформами
Квантовая готовность (Quantum Ready)	Понимающие и умеющие писать квантовые программы, а также разбирающиеся в сложных алгоритмах
Квантовый профессионализм (Quantum Professional)	Имеющие знания и навыки для выхода на рынок

4. Методология

При выполнении обзора литературы мы опирались на рекомендации авторов работ [59-61] по систематическому сопоставлению. Мы также рассмотрели рекомендации для систематических обзоров литературы, данные в работах [62-67], особенно важные для выработки критериев отбора. Иногда для большего охвата во внимание принимаются дополнительные источники [66], используется ручной и менее структурированный поиск в Интернете и других источниках [68] и малотиражная (“серая”) литература, как рекомендовано в [69]. В нашем исследовании мы использовали и формально опубликованные, и “серые” источники.

Далее в этом разделе мы описываем процесс систематического отбора источников для последующего анализа, выделяя основные элементы. Полная версия используемой методологии представлена в статье [34].

Формулировка исследуемых вопросов (Результат: рамки исследования). Решающее значение для вычислений имеют методы и средства программирования [30], при этом, благодаря способности квантовых вычислений решать сложные задачи в различных областях, набирает популярность квантовое программирование [32, 70-72]. Квантовое программирование важно уже потому, что именно оно позволяет использовать все возможности квантовых систем, позволяя решать ранее неразрешимые задачи [31, 73]. Важно и то, что для создания квантового программного обеспечения требуются дополнительные знания и навыки.

Цель обзора литературы: Цель состоит в том, чтобы определить навыки, необходимые в квантовой индустрии, особенно для реализации квантовых вычислений и разработки квантового программного обеспечения. Это включает в себя определение необходимых квантовых знаний и признание как технических (жестких), так и мягких навыков.

Важно отметить, что в предыдущем исследовании [74] представлено первое предложение фундаментальных концепций, связанных с квантовыми вычислениями. В настоящем исследовании набор фундаментальных концепций расширен.

Цель обзора литературы: Цель состоит в том, чтобы определить навыки, необходимые в квантовой индустрии, особенно при проведении квантовых вычислений и разработке квантового программного обеспечения. Поиск направлен и на выявление необходимых

квантовых знаний и признание как технических (жестких, hard), так и социальных (мягких, soft) навыков.

Важно отметить, что однажды уже проводилось исследование [74], в котором было получено первичное представление о фундаментальных концепциях, лежащих в основе квантовых вычислений. В настоящем исследовании этот набор фундаментальных концепций расширен.

Вопросы исследования: Были поставлены три основных вопроса:

RQ1: *Какими знаниями должны обладать сотрудники, выполняющие квантовые вычисления?*

RQ2: *Какими техническими навыками должны обладать сотрудники, выполняющие квантовые вычисления?*

RQ3: *Какими социальными навыками должны обладать сотрудники, выполняющие квантовые вычисления?*

Поиск фраз: Для поиска использовалась следующая фраза:

("quantum computing")
AND ("software engineer" OR "software developer" OR programmer)
AND (skills OR competences)

Рассматривавшиеся базы данных: Рассматривались четыре базы данных: ACM, IEEE Xplore, Scopus, и ScienceDirect.

Период: Рассматривались источники, опубликованные до июня 2023 года, когда проведение исследования было завершено.

Отбор данных для первичного анализа (Результат: отобраны все документы).

Выбор релевантных документов: Результат поиска показан в табл. 3. Из четырех баз данных по поисковой фразе отобраны более 400 элементов.

Табл. 3. Исследовавшиеся базы данных.
Table 3. Databases consulted.

База данных	**Число обнаруженных статей (разделов)**
ACM (AC)	232
ScienceDirect (SD)	105
Scopus	69
IEEE Xplore (XP)	32

Просмотр документов на предмет их включения в анализ или исключения из него (Результат: релевантные документы).

Отбор наиболее важных целей: Мы определили основу для идентификации полученных документов: демографические аспекты, критерии включения и исключения.

Демография: Определяется характер документа.

D1 – Тип документа: исследовательские работы и другие типы документов со значительным вкладом.

D2 – Происхождение документа: журналы, конференции и другие источники.

D3 – Язык: рассматриваются только работы, написанные на английском языке.

D4 – Доступность: полные тексты должны быть доступны через институциональные учетные записи.

Критерии включения: Мы установили двухуровневые критерии включения, имеющие вид фильтров.

Уровень 1: Заголовок документа.

F1: Название включает значимые ключевые слова или фразы, связанные с поисковой фразой.

F2: Раздел ключевых слов содержит важные слова или фрагменты поисковой фразы.

F3: Аннотация устанавливает четкую взаимосвязь между “квантовыми вычислениями”, “разработкой программного обеспечения” и “навыками”.

Уровень 2: Основной текст статьи.

B1: Результаты, которые вводят и описывают концепции квантовых вычислений и/или квантовых программ, а также необходимые навыки для их реализации.

B2: Результаты, в которых со ссылками на необходимые навыки обсуждаются квантовые вычисления и некоторые аспекты квантового программного обеспечения, или наоборот.

Критерии исключения: Критерии исключения применяются для удаления документов, не в которых не обнаруживается нужная информация, данные которых неполны или недоступны для исследования.

EX01: Удалить дубликаты, найденные в базах данных (крайне редкая ситуация).

EX02: Удалить документы, не являющиеся исследовательскими работами (если есть другие работы со значительным вкладом).

EX03: Удалить документы, представляющие собой не полные тексты, а рефераты.

EX04: Удалить документы, написанные не на английском языке (таких не найдено).

EX05: Документы, которые не вводят и не описывают концепции квантовых вычислений и/или квантовой программной инженерии.

EX06: Документы, в которых не обсуждается связь между вычислениями и/или квантовыми программами, инженером-программистом и навыками.

Результаты отбора: Были применены критерии исключения EX01, EX02, EX03 и EX04, а также фильтры уровня 1, полученные результаты отражены в табл. 4. В табл. 5 статьи помечены последовательно пронумерованными обозначениями как S_1, S_2, …S_n.

Табл. 4. Исследованные базы данных.
Table 4. Databases consulted.

База данных	**Число обнаруженных статей (разделов)**
ACM (AC)	37
ScienceDirect (SD)	6
Scopus	5
IEEE Xplore (XP)	3

Ключевые слова в аннотациях (результат: Схема классификации).

В соответствии с целью исследования для классификации отобранных работ были выделены три категории:

1) Требования к навыкам в области программирования (**категория 1**).
2) Образование/учебная программа – развитие навыков квантовых вычислений и практического программирования (**категория 2**).
3) Общие требования к навыкам квантовых вычислений (**категория 3**).

Категория 1: В эту категорию входят самые общие требования, предъявляемые к работникам компаний, выполняющим разработку программного обеспечения; сюда же относятся требования к программистам, работающим в этих компаниях.

Категория 2: В эту категорию входят требования по развитию навыков квантовых вычислений и разработке программного обеспечения, обеспечению связи между обеими областями.

Категория 3: К этой категории относятся требования к общим навыкам, необходимым для квантовых вычислений, особое внимание уделяется пониманию природы квантовых вычислений и конкретным навыкам, которых они требуют.

Табл. 5. Отобранные статьи после применения критериев включения и исключения.
Table 5. Selected papers after applying all the inclusion and exclusion criteria -systematic mapping.

ID_DB	ID_TR	REF.	**Заголовок статьи**
ACM:			
ACM-2	S_1	[75]	Teaching Quantum Computing through a Practical Software-driven Approach: Experience Report
ACM-3	S_2	[3]	Quantum Computing: A New Software Engineering Golden Age
ACM-4	S_3	[33]	Towards Higher-Level Abstractions for Quantum Computing
ACM-5	S_4	[35]	Quantum Software: Model-driven or Search-driven? A Q-SE 2021 Workshop Report
ACM-6	S_5	[1]	Quantum computing: challenges and opportunities
ACM-7	S_6	[25]	Bugs in Quantum computing platforms: an empirical study
ACM-8	S_7	[26]	Making Quantum Computing Open: Lessons from Open-Source Projects
ACM-9	S_8	[23]	Software Engineering Education of Classical Computing vs. Quantum Computing: A Competency-Centric Approach
ACM-10	S_9	[76]	The business of quantum computing
ACM-12	S_{10}	[77]	Quantum Computing as a Topic in Computer Science Education
ACM-15	S_{11}	[78]	First International Workshop on Quantum Software Engineering (Q-SE 2020)
ACM-18	S_{12}	[79]	Entanglion: A Board Game for Teaching the Principles of Quantum Computing
ACM-20	S_{13}	[80]	QAI4ASE: Quantum artificial intelligence for automotive software engineering
ACM-21	S_{14}	[81]	Hybrid quantum-classical problem solving in the NISQ era
ACM-22	S_{15}	[82]	Thinking Too Classically: Research Topics in Human-Quantum Computer Interaction
ACM-23	S_{16}	[83]	Building a quantum computer
ACM-25	S_{17}	[84]	Exploring Quantum Reversibility with Young Learners
ACM-26	S_{18}	[85]	The Holy Grail of Quantum Artificial Intelligence: Major Challenges in Accelerating the Machine Learning Pipeline
ACM-27	S_{19}	[86]	Cyberinfrastructure Facilitation Skills Training via the Virtual Residency Program
ACM-31	S_{20}	[87]	Reflections on cyberethics education for millennial software engineers
ACM-32	S_{21}	[88]	Qupcakery: A Puzzle Game that Introduces Quantum Gates to Young Learners
ACM-39	S_{22}	[89]	65 competencies: which ones should your data analytics experts have?
ACM-40	S_{23}	[90]	Interactive online tool as an instrument for learning mathematics through programming techniques, aimed at high school students
ACM-41	S_{24}	[91]	Evolution of the Competencies to Embrace Digital Technology for Sustainable Development
ACM-46	S_{25}	[92]	A golden age for computing frontiers, a dark age for computing education?
ACM-47	S_{26}	[93]	Beyond the badge: reproducibility engineering as a lifetime skill
ACM-48	S_{27}	[94]	To Write Code: The Cultural Fabrication of Programming Notation and Practice
ACM-49	S_{28}	[95]	Should Quantum Processor Design be Considered a Topic in Computer Architecture Education?
IEE Xplore:			
XP-07_	S_{38}	[96]	Statistical Assertions for Validating Patterns and Finding Bugs in Quantum Programs
XP-12	S_{39}	[97]	A Software Development Kit and Translation Layer for Executing Intel 8080 Assembler on a Quantum Computer
XP-13	S_{40}	[98]	Need and Challenges in Quantum Computing in Fog Environment
Scopus:			
SC-7	S_{41}	[99]	Quantum Programming on Azure Quantum—An Open Source Tool for Quantum Developers
SC-21	S_{42}	[100]	Society 5.0 and the future of work skills for software engineers and developers
SC-24	S_{43}	[101]	Industry quantum computing applications
SC-52	S_{44}	[102]	New Programming Paradigms.
SC-57	S_{45}	[103]	Introducing Microsoft Quantum Computing for Developers: Using the Quantum Development Kit and Q#
ScienceDirect:			
SD-4	S_{46}	[104]	Reigniting the power of artificial intelligence in education sector for the educators and students' competence.
SD-6	S_{47}	[105]	QuantuMoonLight: A low-code platform to experiment with quantum machine learning
SD-20	S_{48}	[106]	Industry 4.0 in Terms of Industrial Relations and Its Impacts on Labour Life
SD-26	S_{49}	[107]	The future of sustainable digital infrastructures: A landscape of solutions, adoption factors, impediments, open problems, and scenarios
SD-33	S_{50}	[108]	Industry 4.0 in a project context: Introducing 3D printing in construction projects
SD-55	S_{51}	[109]	Fresh Outlook on Numerical Methods for Geodynamics. Part 2: Big Data, HPC, Education

Извлечение данных и их отнесение к той или иной категории (результат: Систематизация результатов).

В табл. 6 показан результат группирования документов по трем указанным категориям. Как легко видеть, наибольшее число статей попало в категорию 2, следом идет категория 3, а категория 1 по размерам близка к категории 3. Результаты показывают, что учебные заведения серьезно относятся к подготовке квалифицированных выпускников и стремятся соответствовать требованиям промышленности и вызовам, предъявляемым со стороны квантовых вычислений.

Табл. 6. Распределение работ по категориям.
Table 6. Grouping of papers into categories.

Категория	Источники
1. Требования к квалификации для работы в отрасли	S_{13}, S_{18}, S_{22}, S_{24}, S_{26}, S_{41}, S_{42}, S_{43}, S_{44}, S_{45}, S_{50}
2. Образование/учебная программа (развитие навыков квантовых вычислений и программной инженерии)	S_1, S_4, S_8, S_{10}, S_{11}, S_{12}, S_{13}, S_{14}, S_{17}, S_{19}, S_{20}, S_{21}, S_{23}, S_{25}, S_{26}, S_{27}, S_{28}, S_{29}, S_{30}, S_{31}, S_{32}, S_{33}, S_{34}, S_{35}, S_{36}, S_{37}, S_{44}, S_{46}, S_{49}, S_{51}
3. Общие требования к навыкам по квантовым вычислениям	S_2, S_3, S_5, S_6, S_7, S_9, S_{15}, S_{16}, S_{18}, S_{38}, S_{39}, S_{40}, S_{45}, S_{47}

5. Результаты

Отвечая на поставленные исследовательские вопросы, подробно рассмотрим три основных компонента компетенции – знания, технические навыки, социальные навыки.

RQ1: Какими знаниями должны обладать сотрудники, выполняющие квантовые вычисления?

Чтобы ответить на этот вопрос, мы рассмотрели основные квантовые концепции, упоминаемые в изученных источниках из табл. 6. При этом, чтобы иметь полное понимание основных концепций, некоторые определения, представленные в этом разделе, были взяты из вспомогательной литературы. В качестве основы мы рассматриваем фундаментальные концепции квантовых вычислений, предложенные в работе [110].

Будем отталкиваться от таксономии, строящейся с учетом самых необходимых знаний. Это позволит нам получить более полную, всеобъемлющую базу знаний.

Квантовая теория информации (QIT). Эта дисциплина охватывает значительный набор знаний. Ее задача – исследование достижимых пределов обработки информации в рамках квантовой механики [111]. На рис. 1 показана связь этой науки с другими научными областями.

Квантовая информация – это информация о состоянии квантовой системы. Она является основным объектом исследования в квантовой теории информации [112], квантовую информацию можно обрабатывать с помощью методов квантовой теории информации.

Теория информации. Теория информации – это математический подход к оценке количества информации, к изучению методов хранения и передачи информации [113] (см. рис. 2).

К важным разделам теории информации относятся кодирование алфавитов символов, теория алгоритмической сложности, алгоритмическая теория информации и теоретико-информационная безопасность (см. рис. 3).

Квантовая механика. Квантовая механика – область физики, изучающая поведение элементарных частиц на микроскопическом уровне [114-115]. Основные понятия квантовой механики показаны на рис. 4.

Рис. 1. Квантовая информатика как многодисциплинарная композиция.
Fig. 1. Quantum Information Science -interdisciplinary composition.

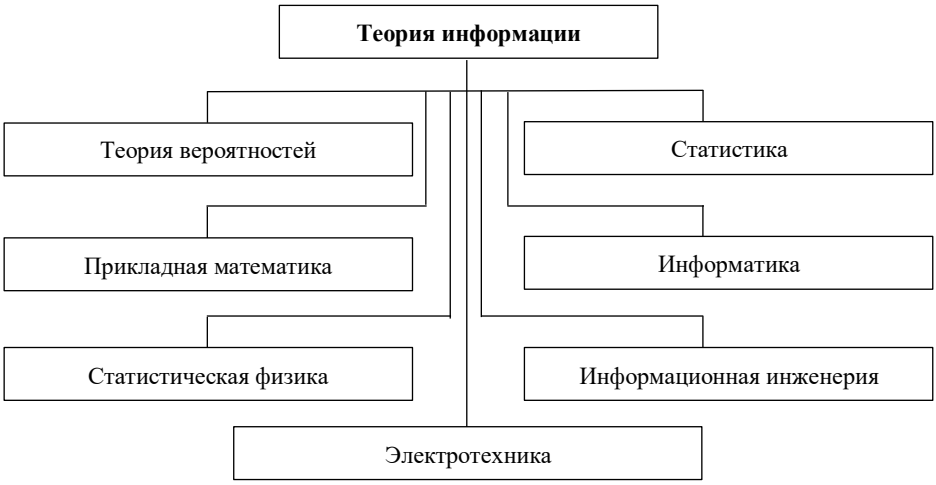


Рис. 2. Теория информации как пересечение научных дисциплин.
Fig. 2. Information Theory -interdisciplinary intersection.

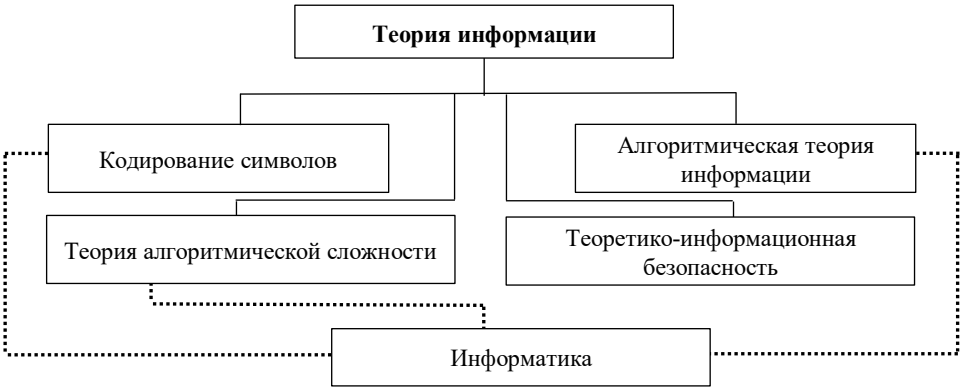


Рис. 3. Составные части теории информации.
Fig. 3. Information Theory -subfields.

Рис. 4. Базовые концепции квантовой механики.
Fig. 4. Quantum Mechanics – fundamental concepts.

Хорошо описанные определения основ квантовой механики, такие как суперпозиция, запутанность и декогерентность, а также кубит как фундаментальная единица вычисления, представлены в работах [115-117] и подробно объяснены в работе [118].

Основные понятия квантовой механики проиллюстрированы на рис. 5, где с помощью конкретных понятий обозначены некоторые основополагающие концепции.

Математика. Математика как наука связана с квантовыми вычислениями по двум направлениям – использованием вероятностного подхода при измерении квантовых состояний и применением методов линейной алгебры для матричного представления квантовых вентилей, являющихся физической основой выполнения квантовых вычислений. Математические концепции, используемые в квантовой механике, показаны на рис. 6.

Физика. Физика является фундаментальной наукой, которая, кроме прочего, включает в себя и квантовую механику [119-120], поэтому так важно изучать фундаментальные концепции этой науки. В работах [121-123] подчеркивается вклад физики в квантовые вычисления. На рис. 7 представлен список тем для рассмотрения.

Информатика. Наука информатика состоит в изучении методов вычислений, информации и автоматизации [124-126] (см. рис. 8). Квантовые вычисления своими важнейшими концепциями и принципами опираются на информатику [10]. Понятия, вводимые в информатике, которые могут иметь отношение к квантовым вычислениям, показаны на рис. 9.

RQ2: Какими техническими навыками должны обладать сотрудники, выполняющие квантовые вычисления?

Сложные навыки, называемые техническими, – это любые навыки, относящиеся к конкретной задаче или ситуации. Они включают в себя как понимание, так и владение специфической деятельностью, включающей методы, способы, процессы, процедуры [127]. Технические навыки тесно связаны со знаниями, основу знаний для технических навыков создает ответ на вопрос RQ1. Чтобы ответить вопрос RQ2, мы дополнительно рассмотрели наиболее распространенные технические навыки, указанные в авторитетных источниках перечисленных в табл. 6. При этом из вспомогательной литературы были извлечены даже более комплексные предложения. В работе [128] представлен список навыков, необходимых в квантовой индустрии, разделенный на группы по рекомендуемым курсам обучения. Авторы этой работы предложили формат интервью, которое помогает уточнить необходимые навыки. Списки научных и технических навыков, использованные в качестве подсказок в

протоколе собеседования, были взяты из отчета Объединенной целевой группы по программам бакалавриата по физике за 2016 год [123]. В этой вспомогательной работе особое внимание уделяется знаниям и научному опыту.

Рис. 5. Квантовая механика – детализированный взгляд.
Fig. 5. Quantum Mechanics -a detailed view.

Знания, специфичные для физики: Цели обучения физике включают в себя способность применять фундаментальные концепции, в частности, законы сохранения, для решения задач, компетентность в использовании базовых законов физики в различных областях и контекстах, а также умение представлять эти концепции несколькими способами при решении междисциплинарных проблем.

Научно-технические навыки: Цели обучения научно-техническим навыкам включают решение сложных задач с проведением экспериментов, моделирования и построением аналитических моделей; планирование дальнейших исследований и определение потребностей в ресурсах. Они также включают в себя умение работать с приборами, проектировать и писать программы, кодированием и проводить анализ данных.

За основу для распределения результатов по категориям нами был взят именно такой набор технических (жестких) навыков. Табл. 7 демонстрирует общий взгляд на четыре выделенные нами категории навыков. Описание навыков, необходимых для квантовых вычислений, по каждой категории в этой таблице дано практиками-специалистами, здесь же указана литература, описывающая эти навыки. Табл. 8 содержит набор навыков, необходимых для

выполнения реальных работ и связанных с теорией квантовой информации. Табл. 9, табл. 10 и табл. 11 содержат наборы навыков, рекомендованных для конкретных учебных курсов.

Описание, данное для каждого навыка, является практической рекомендацией для тех, кто в своей работе должен демонстрировать свою способность применять знания и навыки при выполнении рабочих заданий, связанных с квантовыми вычислениями.

RQ3: Какими социальными навыками должны обладать сотрудники, выполняющие квантовые вычисления?

Ученые, работающие в области квантовой информации, опираются на работу физиков, математиков, специалистов, развивающих теорию и практику программирования, а также материаловедов, химиков и инженеров [18]. Междисциплинарные команды применяют квантовую информатику, развивая новые квантовые технологии в области связи, информационных сетей, безопасности данных, навигации и медицинской диагностики (см. рис. 10). Кроме этого, для эффективной командной работы и решения поставленных перед ними задач практические специалисты квантовых вычислений нуждаются в значительных навыках другого рода. Чтобы контролировать ведущиеся проекты, менеджеры квантовых проектов должны обладать глубокими знаниями процессов, способов организации, принципов, политик, информации, культуры, этики, командных навыков и компетенций команды, а также связанных с работами услуг, инфраструктуры и приложений.

Такие “мягкие” навыки, которые мы здесь называем социальными, – это сочетание межличностных, человеческих навыков, навыков общения, понимания черт характера, отношений, карьерных соображений, уровня эмоционального интеллекта и многого другого [129]. Основываясь на этом определении, мы собрали основные социальные навыки, упоминаемые в рассмотренной нами литературе (см. табл. 12). Эти навыки горизонтальны; ими должны владеть и демонстрировать профессионалы в любой дисциплине. Например, в работе [123] представлен отчет для программ физики, выделяющий навыки, требуемые от промышленности. В этой работе авторы выделили две группы навыков, а именно:

Коммуникативные навыки: Цели обучения коммуникативным навыкам включают способность общаться устно и письменно с аудиториями, которые отличаются широким диапазоном приобретенного опыта и наличных потребностей.

Профессиональные навыки и навыки, применяемые на рабочем месте: Цели обучения профессиональным навыкам и навыкам на рабочем месте включают работу в различных командах, понимание свойств соответствующих технологических ресурсов и знакомство с концепциями рабочего места, такими как управление проектами, бюджетирование, оценка качества, а также нормативные и регулятивные вопросы.

В табл. 11 мы собрали литературные источники, систематизацию которых мы проводили и в которых мы встречали описание тех или иных навыков. Кроме того, мы включили в нее некоторые дополнительные использованные нами источники. Необходимо подчеркнуть, что все лидеры отрасли говорили о важности иметь возможность добиться успеха на командных проектах и эффективно доносить технические идеи до широкой аудитории [130].

Социальные навыки обычно приобретаются благодаря опыту, хотя некоторые из них могут быть развиты с помощью учебных программ. Выпускники бакалавриата с профессиональным опытом могут иметь важное значение для ролей, требующих базовых квантовых знаний, в то время как обладатели степени PhD со специализированной квантовой подготовкой часто при изучении вакансий сталкиваются с заниженными требованиями к профессиональному опыту. В работе [131] утверждается, что академические программы должны предлагать несколько точек входа для построения карьеры в квантовой индустрии – для тех, кто еще учится, для обладателей дипломов бакалавра и магистра, для прошедших производственную практику и закончивших аспирантуру. Сочетание традиционного образования с практическим

обучением и участием в реальных проектах будет иметь решающее значение для развития необходимых навыков для востребованных рабочих мест.

Рис. 6. Математика – детализированный перечень концептов для квантовых вычислений.
Fig. 6. Mathematics -detailed list of concepts for quantum computing.

Рис. 7. Физика – детализированный взгляд на фундаментальные концепции.
Fig. 7. Physics -detailed view of fundamental concepts converging to quantum mechanics.

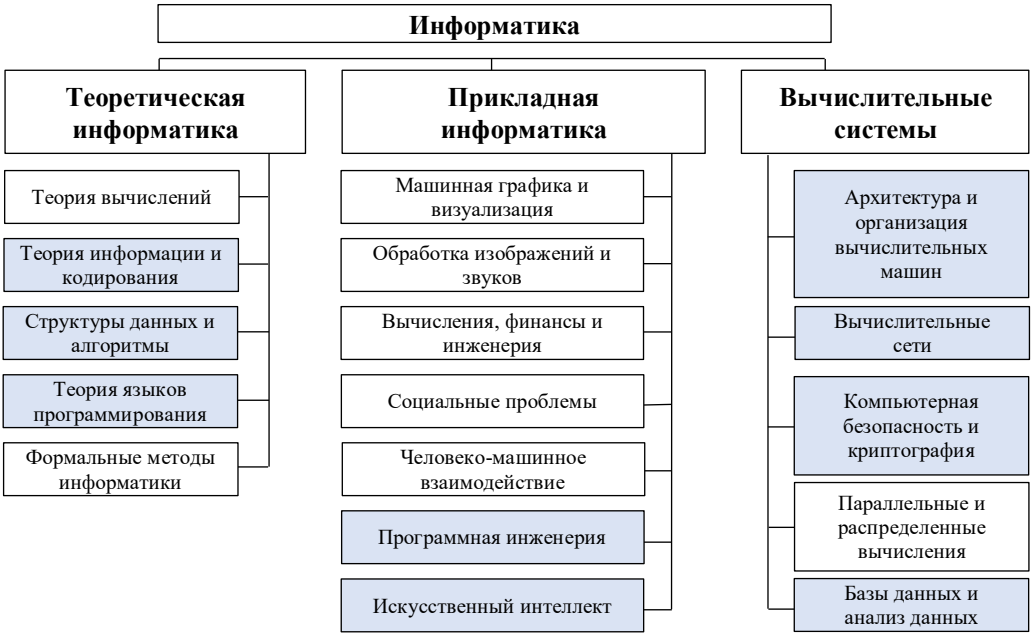


Рис. 8. Информатика – детализированный взгляд на составляющие и смежные дисциплины.
Fig. 8. Computer Science -detailed view of subfields and related disciplines.

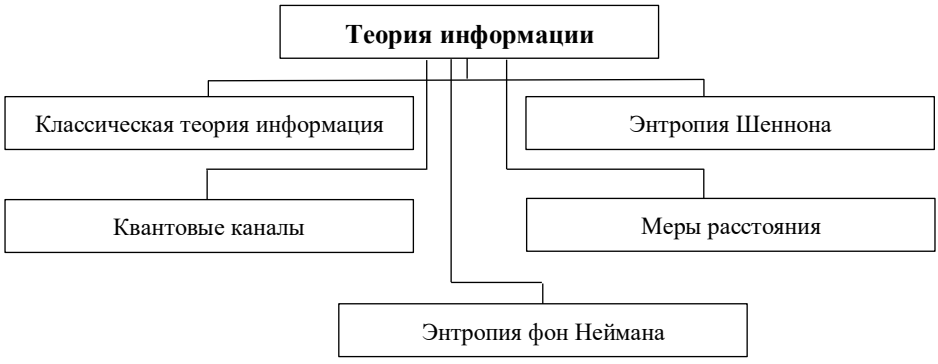


Рис. 9. Информатика – специфическая тематика.
Fig. 9. Computer Science -specific topics.

6. Обсуждение

Основные выводы:

- **Расширение сферы применения:** Наблюдается рост различных областей применения квантовых вычислений, рабочие места заполняются специалистами с образованием в области физики.
- **Охват знаний:** Изучение литературы показывает обширное цитирование квантовых знаний, охватывающее основные понятия из квантовой информатики и связанных дисциплин, таких как квантовая механика, математика и информатика.
- **Образовательные усилия:** Различные научные и образовательные инициативы направлены на подготовку студентов и специалистов к использованию квантовых вычислений, хотя единого мнения о содержании и глубине курсов еще нет.

Табл. 7. Классификация практических навыков в области квантовых вычислений.
Table 7. Classification of skills for quantum workforce.

Категория	Навыки
Владение теорией квантовой информации для обычного мира	Кодирование; Статистический анализ и анализ данных; Поиск неисправностей (отладка); Источники шума; Моделирование; (Де)когерентность; Исправление ошибок; Динамика открытых систем; Оборудование для кубитов; Гамильтонианы; Проектирование квантовых цепей (физическое).
Навыки квантовой индустрии для возможного курса по традиционной квантовой теории	Линейная алгебра; Статистика (в отношении к измерениям); Гамильтонианы уравнения Шредингера унитарной временно́й эволюции; Гильбертово пространство; Запутанность; Двухуровневые системы; Спектроскопия; Атомные энергетические уровни и фотонное взаимодействие.
Навыки квантовой индустрии для возможного курса традиционной квантовой теории информации	Линейная алгебра; Статистика (в отношении к измерениям); Источники шума; Проектирование квантовых цепей и вентильных моделей; Алгоритмы; Абстрактные кубиты; Теория сложности; Квантовые безопасные коммуникации; Гамильтонианы; Высокопроизводительные вычисления.
Навыки, полезные на рабочих местах квантовой индустрии, для которых рекомендуется читать образовательные курсы по квантовой информатике	Программное обеспечение; Кодирование; Лабораторные навыки и лабораторный опыт; Устранение неисправностей; Материалы; Холодные атомы; Ионные ловушки; Лазерное охлаждение; Магнитооптические ловушки; Конденсаты Бозе-Эйнштейна; Сверхпроводящие цепи; Фотонные интегральные схемы; Высокопроизводительные вычисления; Отжиг.

Табл. 8. Владение теорией квантовой информации для обычного мира.
Table 8. Skills Real-world Quantum Information Theory.

Навык	Описание	Основные источники
Кодирование	Уметь программировать с использованием традиционных алгоритмов, уметь отлаживать программы на квантовом процессоре, уметь собирать и анализировать данные и обсуждать результаты с программистами.	S_6, S_7, S_8, S_{15}, S_{23}, S_{26}, S_{27}, S_{39}, S_{45}
Статистический анализ и анализ данных	Уметь уточнять процессы и калибровки компьютера, который участвует в экспериментах, и понимать, какие гипотезы были отвергнуты, а какие доказаны.	S_{23}, S_{29}
Устранение неисправностей (отладка)	Уметь выполнять отладку, радикально отличную от классического варианта, поскольку в классических отладчиках программист имеет возможность расставлять точки останова, а в квантовых программах делать этого нельзя, так как за этим скрывается измерение, разрушающее информацию.	S_{15}, S_{41}, S_{45}
Источники шума	Уметь управлять механизмами шумов. Какой вентильный автомат используется, на каких принципах он работает, понимать механику декогерентности и шума, а также механику обычных сбоев для обеих моделей квантовых вычислений.	S_{16}, S_{51}
Моделирование	Быть способным понимать, как получить зашумленные данные, понимать, что находится в основе физической модели, в соответствии с которой переносятся эти данные, знать, как очистить эти данные и делать статистически значимые выводы о правильности предложенной модели или о том, как ее улучшить.	S_2, S_3, S_4, S_8, S_{14}, S_{26}
(Де) когерентность	Уметь управлять механизмом декогерентности, знать технологию, опираясь на которую оперирует датчик. Для конкретного приложения важно знать, нужна ли в данном случае система с более длительным временем когерентности или система, которую можно легко связать с какой-либо другой системой. Каковы пределы этого связывания?	S_9, S_{10}, S_{11}, S_{12}, S_{16}, S_{17}, S_{28}, S_{29}
Исправление ошибок	Уметь управлять квантовой коррекцией ошибок. Для квантовой коррекции ошибок, проектирования оборудования и моделирования необходимо много знать о физике процессов.	S_{11}, S_{12}, S_{16}, S_{17}
Динамика открытых систем	Понимать динамику открытой системы и управлять ею. Уравнение Шредингера определяет эволюцию во времени, которая даже в большей степени, чем унитарная динамика, позволяет увидеть и, можно надеяться, понять динамику открытых систем.	S_4, S_7
Оборудование для кубитов	Понимать поведение кубитов, особенно, как (в терминах механизмов декогерентности) распространяется информация.	S_{14}, S_{16}, S_{45}
Гамильтонианы	Уметь использовать мощности оператора Гамильтона. Связь абстрактного Гамильтониана с актуальной физической реальностью, конечно же, не очевидна, и ее ограничения нуждаются в тщательном понимании.	S_4, S_{43}
Проектирование квантовых цепей (физическое)	Уметь проектировать квантовые цепи: взяв спецификацию квантовой системы в терминах Гамильтониана, превращать ее в проект для изготовления. Здесь есть место для многих микроволновых инженерных решений, и одновременно видна тесная связь с квантованием микроволновых цепей.	S_3, S_{14}, S_{16}

Табл. 9. Навыки квантовой индустрии для возможного курса по традиционной квантовой теории.
Table 9. Skills of relevance to the quantum industry for a possible traditional quantum theory course.

Навык	Описание	Дополнительные источники
Линейная алгебра	Уметь на абстрактном уровне использовать/применять линейную алгебру, необходимую для квантовой теории, включая абстрактные векторные пространства, абстрактные операторы, комплексные векторные пространства.	S_{23}, S_{29}
Статистика (в отношении к измерениям)	Уметь строить статистики для измерений. Это необходимо, так как взаимодействие фотонов с атомами имеет статистическую природу.	S_7, S_{23}, S_{29}
Гамильтонианы уравнения Шредингера унитарной временно́й эволюции	Уметь управлять унитарной временно́й эволюцией. Поскольку временна́я эволюция описывается уравнением Шредингера, необходимо в достаточной степени обладать пониманием Гамильтонианов и временно́й эволюции.	S_4, S_7, S_{43}
Гильбертово пространство	Уметь работать с Гильбертовым пространством. Каждый член команды участников проекта должен знать, как именно Гильбертово пространство, унитарная временна́я эволюция, измерения и запутанность связаны с базовыми формализмами квантовой механики.	S_4, S_7, S_{43}
Запутанность	Уметь управлять квантовой запутанностью, как одним из наиболее важных понятий.	S_4, S_7, S_9, S_{10}, S_{11}, S_{12}, S_{16}, S_{17}, S_{43}, S_{47}
Двухуровневые системы	Уметь проектировать двух- и трехуровневые системы, а также разбираться в уравнениях скорости для оптической накачки и истощения.	S_4, S_7, S_{43}
Спектроскопия	Быть способным вести полезные обсуждения атомных спектров, делать обоснованные утверждения, наподобие таких, как “*это не будет работать, так как правила отбора запрещают переход*”.	S_4, S_7, S_{43}
Атомные энергетические уровни и фотонное взаимодействие	Понимать, что такое атомные энергетические уровни и фотонное взаимодействие, уметь управлять этими явлениями, то есть иметь знания в области атомной физики, достаточные для вычисления уровней, моделирования и для понимания факторов Франка–Кондона.	S_4, S_7, S_{43}

- **Основные темы:** Хотя согласия по конкретным темам для курсов по квантовым вычислениям нет, имеются предложения по ключевым темам, виден и некоторый успех в преподавании на старших курсах и в магистратуре, в то время как потребности индустрии в основном поддерживаются аспирантскими программами.
- **Важность социальных навыков:** В литературе подчеркивается важность социальных навыков из-за междисциплинарного характера квантовых вычислений, подчеркивая этим необходимость эффективной практической работы с разнородными командами.

Выводы по проблемным вопросам:

- **Отсутствие отраслевых стандартов:** В отрасли не выработаны стандарты для определения требований к работе в квантовых вычислениях, а необходимые навыки часто не указываются явно.
- **Отсутствие академического консенсуса:** Нет согласия по темам для курсов квантовых вычислений, что подчеркивает настоятельную необходимость определения ключевых знаний для разработки структурированных, воспроизводимых курсов.
- **Направленность преподавания:** Большинство усилий в области преподавания направлены на программы для выпускников полного курса, но выпускники бакалавриата также востребованы для исполнения технической работы в квантовых вычислениях.

Табл. 10. Навыки квантовой индустрии для возможного курса по традиционной квантовой теории. Table 10. Skills of relevance to the quantum industry for a possible quantum information theory course.

Навык	**Описание**	**Дополнит. источники**
Линейная алгебра	Уметь применять линейную алгебру при объяснении квантовых феноменов и линейные операторы для работы с кубитами. Теоретики и практики должны владеть линейной алгеброй, операциями над матрицами, демонстрировать, что происходит в квантовых цепях.	S_4, S_7, S_{23}, S_{29}, S_{43}
Статистика (в отношении к измерениям)	Уметь применять статистические методы для описания квантовых феноменов, рассматривая квантовую механику в качестве обобщенного примера классической теории вероятностей. Выборка из распределения вероятностей системы.	S_7, S_{23}, S_{29}
Источники шума	Уметь работать с зашумленной системой. Таковыми являются квантовые цепи и вентили, а также модели с шумами.	S_{11}, S_{12}, S_{16}, S_{17}, S_{51}
Проектирование квантовых цепей и вентильных моделей	• Понимать и уметь проектировать цепи (свободные от шума) и вентильные модели. Полное моделирование шума, модели вентилей и алгоритмы оптимизации должны быть разделены, так как не все вентили одинаковы. Отсюда следует, что теоретики должны искать новые способы снижения шумов. • Уметь правильно и безошибочно применять все типы вентилей. Не обязательно знать все наизусть, вроде того, какая матрица соответствует вентилю CNOT, а какая вентилю Адамара, но надо уметь показать, что делает система, какое состояние она получит в итоге, как все эти эволюции описываются матрицей-произведением или набором матриц Паули.	S_3, S_{14}, S_{16}
Алгоритмы	• Уметь строить и использовать квантовые алгоритмы. Для практической работы требуется понимать модель данных, понимать, как свести ее к некоторому алгоритму, который может быть полезен для решения конкретной задачи. • На практике не требуется помнить все известные алгоритмы, но нужно знать, какие алгоритмы имеет смысл попробовать. • На практике требуется умение разобраться в квантовом алгоритме (а не просто использовать квантовый язык программирования), а также понимать, можно ли решить поставленную задачу с помощью квантового компьютера.	S_1, S_2, S_{13}, S_{14}, S_{17}, S_{19}, S_{45}, S_{51}
Абстрактные кубиты	Уметь интерпретировать абстрактную реализацию кубита: сфера Блоха – это очень полезная иллюстрация пространства состояний кубита, надо хорошо знать этот подход, понимать язык описания абстрактных кубитов, поскольку этот язык используется постоянно и повсеместно.	S_4, S_7, S_9, S_{10}, S_{11}, S_{12}, S_{14}, S_{16}, S_{17}, S_{43}, S_{45}, S_{47}
Теория сложности	• Владеть теорией сложности: квантовые компьютеры полностью базируются на информатике, ее теоретических положениях, поэтому нужно знать теорию сложности, уметь упорядочивать алгоритмы по времени их работы, понимать, как теория помогает ускорить решение задачи. • Уметь преодолевать рамки классических представлений, понимать, что, например, алгоритм Шора неструктурированного поиска имеет сложность, пропорциональную корню из N, а не N.	S_7, S_{10}, S_{12}, S_{43}
Квантовые безопасные коммуникации	Уметь применять квантовые алгоритмы в безопасных коммуникациях. Что такое алгоритмы шифрования, что такое подпись, как будут реализованы некоторые протоколы безопасности.	S_4, S_{35}, S_{36}, S_{37}, S_{43}
Гамильтонианы	Уметь применять операторы Гамильтона к решению задач: с каждой реализацией сверхпроводящего кубита связан некоторый Гамильтониан и энергетический ландшафт. Надо понимать, какой это Гамильтониан, как получить собственные состояния.	S_4, S_{43}
Высоко-производительные вычисления	Уметь применять методы и приемы для высокопроизводительных приложений. Нужно понимать, какие современные суперкомпьютеры нужны для квантовых вычислений, а также для некоторых прикладных работ. Это, например, методы, основанные на квантовых технологиях, которые в конечном итоге были использованы на высокопроизводительных машинах.	S_{35}, S_{36}, S_{37}

Табл. 11. Навыки, полезные на рабочих местах квантовой индустрии, для которых рекомендуется читать образовательные курсы по квантовой информатике.
Table 11. Skills of relevance to the quantum industry for a possible hardware for quantum information course.

Навык	**Описание**	**Дополнит. источники**
Программное обеспечение	Быть способным продемонстрировать знания и умения вести автоматизированное проектирование систем. Уметь моделировать квантовые системы на платформах Mathematica или MATLAB, или каких-либо других. Уметь не просто писать программы, работать с системами управления версиями, в распределенном окружении, где программистам доступны все результаты их коллег.	S_3, S_4, S_8, S_{26}, S_{41}
Кодирование	Быть способным на основе уже достигнутого результата, понять, как может быть автоматизирован процесс его получения, как сделать некоторые вещи лучше, быстрее и проще. Уметь писать программы: вообще говоря, наличие базовых программистских навыков – это, вероятно, абсолютная необходимость для сотрудников разных специальностей. Уметь взаимодействовать с коллегами при решении программистских задач, то есть при написании программ работать с другими людьми, знать, как совместно писать и ревизовать программ, писать резюме к программам, создавать тестовые наборы. Разбираться в структурах данных и знать, как может выглядеть полезный прикладной интерфейс.	S_6, S_7, S_8, S_{15}, S_{23}, S_{26}, S_{27}, S_{39}
Лабораторные навыки и лабораторный опыт	Уметь проводить эксперименты и реализовывать проекты. Люди, которые покупали дрель в магазине, знают, что перед сверловкой надо наметить центр отверстия. Люди с опытом работы в некотором окружении умеют добиваться большего и работают быстрее, так как знают, что вызывает трудности, а что сделать проще. Люди, которые могут спроектировать экспериментальную программу, могут ее реализовать, подобрать данные, понять, для чего эти данные нужны, а затем решить, что делать дальше.	–
Устранение неисправностей	Уметь решать каждодневные проблемы. *Большая часть ежедневных забот – это борьба с программами, борьба с ошибками и выяснение того, что происходит с холодильником*.	S_{15}, S_{41}
Материалы	Уметь создавать и оперировать новыми материалами: Люди, ведущие исследования, изготавливающие новые материалы, стараются понять, как привнести эти новые материалы в свои приборы, что будет хорошим теплоизолятором, что хорошо проводит тепло, что имеет хорошую электропроводность, как эти материалы работают друг с другом, какова теплоемкость тех или иных веществ.	S_4, S_7, S_{43}
Холодные атомы	Знать о лазерах, длинах волн, оптических компонентах, управлении, рассеянном свете и методах измерений.	S_4, S_7, S_{43}
Ионные ловушки	Понимать, что такое технология ионных ловушек, взаимодействие ионов с электроникой, физика самого процесса улавливания ионов, например, уравнения Матье, области устойчивости.	S_4, S_7, S_{43}
Лазерное охлаждение	Понимать и уметь применять материал для лазерного охлаждения, использовать свет для охлаждения атомов, оптических волокон, различных детекторов, электрооптики, акустической оптики.	–
Магнито-оптические ловушки	Уметь работать с материалами магнитооптических ловушек. Построение таких ловушек и улавливающих атомов, опыт, полученный при этом, более ценен, чем получение производных материалов.	S_4, S_7, S_{43}
Конденсаты Бозе-Эйнштейна	Уметь работать с соответствующими материалами и объяснять их поведение. Если кто-то помещает некоторый конденсат в ловушку, которая его детектирует с некоторым фазовым сдвигом или еще как-то, нужно понимать теорию процесса, а тот, кто выполнил такую работу, уметь сделать доклад с объяснением эффекта.	S_4, S_7, S_{43}
Сверхпроводящие цепи	Уметь манипулировать и понимать полупроводниковые цепи. Если кто-то исследует время когеренции или нечто похожее, ему/ей необходимо получить множество сведений по материаловедению и проектированию квантовых цепей и понять мельчайшие детали того, как изготавливается сверхпроводящий кубит, где находятся поля и что имеет значение для когерентных связей.	S_3, S_{14}, S_{16}
Фотонные интегральные схемы	Уметь манипулировать и понимать фотонные интегральные схемы. Чтобы этого добиться, нужно уметь строить кольцевые резонаторы; знать, как работать со стандартными фотонными интегральными схемами, например, в литейном цехе.	S_3, S_4, S_7, S_{14}, S_{16}, S_{43}
Высокопроизводительные вычисления	Быть способным демонстрировать экспертные знания высокопроизводительных вычислений, поскольку в настоящее время множество инженерных задач решается с применением компьютерного моделирования, а значит, при проектировании оборудования необходимо строить программные модели.	S_{35}, S_{36}, S_{37}
Отжиг	Знать, что такое (квантовый) отжиг и принципы, на которых он работает.	S_4, S_7, S_{43}

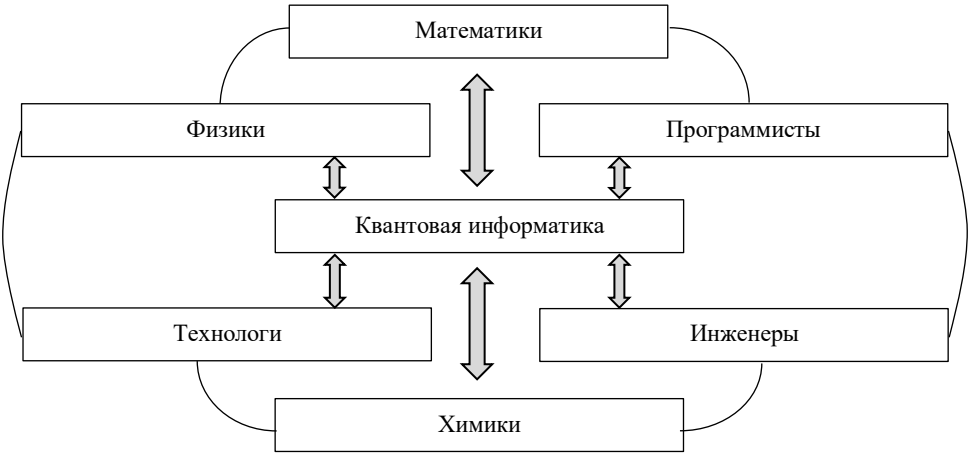


Рис. 10. Междисциплинарные команды специалистов по квантовым вычислениям.
Fig. 10. Cross-disciplinary teamwork of quantum computing practitioners.

Табл. 12. Основные социальные навыки, необходимые для индустрии квантовых вычислений.
Table 12. Main soft skills required for the quantum computing industry.

Социальный навык	**Источники**
Аналитическое/критическое/творческое мышление	S_3, S_{20}, S_{22}, S_{23}, S_{33}, S_{34}
Взаимоотношения	S_{20}, S_{22}, S_{26}, S_{34}, S_{42}
Умение управлять временем	S_{22}, S_{33}, S_{34}, [123]
Командная работа	S_{22}, S_{33}, S_{34}, [123]
Устное/письменное взаимодействие с людьми	S_{22}, S_{33}, S_{34}, [123]
Активное обучение/любопытство	S_{22}, S_{34}, [123]
Самостоятельность, сопряженная с пониманием, когда надо обратиться за помощью	S_{22}, S_{33}, S_{34}, [123]
Способность к междисциплинарной работе	S_{22}, S_{33}, S_{34}, [123]
Умение справляться с двусмысленностями	S_{22}, S_{23}, S_{42}
Лидерство и социальное влияние	S_{22}, S_{23}, S_{33}, S_{34}

7. Заключение

Квантовые вычисления – это новая развивающаяся область, которая может кардинально изменить наши представления о вычислениях и программировании. Сложность, унаследованная от квантовой механики, затрудняет занятия квантовыми вычислениями профессионалам, не знающим физику [73].

Важнейшей задачей развития возможностей персонала является обучение различных групп не проходивших специальный отбор студентов колледжей и университетов навыкам работы с квантовыми вычислениями, для которых которые необходим существенный теоретический базис [132].

На нынешнем этапе второй квантовой революции работа в компаниях-разработчиках и разработка квантовых компьютеров все еще, как правило, требует научного багажа подготовки на уровне подготовки кандидата наук (степени PhD). Однако растущее предложение «квантово-близких» рабочих мест в отрасли также требует квантовой грамотности. Запросы требуют знаний программирования, умения вести разработку программного обеспечения, алгоритмов, владения знаниями электроники, криогеники и

вакуумных технологий [38]. Исходя из этого, для повышения квалификации сотрудников до уровня, необходимого индустрии квантовых вычислений, требуется разработать учебные планы и создать образовательные программы бакалавриата в области квантовых информационных технологий.

Несмотря на некоторые инициативы по созданию учебных пособий и образовательных программ, интеграция квантовых вычислений в информатику, инженерию и связанные с ними учебные планы остается сложной задачей. До сих пор нет консенсуса по тематикам курсов и их последовательности, необходимым для подготовки будущих специалистов.

Пытаясь внести свой вклад в решение назревших задач, в этом отчете мы представляем обзор литературы, вносящий в нее некоторую систематизацию, которая позволила нам выделить следующие вопросы:

1) Полный набор знаний, связанных с квантовыми вычислениями, включая фундаментальные основы физики, квантовой механики, квантовой информатики, математики, информатики и смежных дисциплин.
2) Набор технических навыков (связанных с фундаментальными понятиями квантовых вычислений) с описанием конкретных способностей, связанных с компетенциями участников квантовых вычислений. Эти навыки называются техническими или “жесткими” навыками.
3) Набор “мягких” навыков, называемых в этой статье “социальными”, представляющих собой навыки, требуемые при производстве программного обеспечения, к которым также относятся некоторые конкретные навыки, требуемые самим существом квантовых вычислений.

Будущая работа. Сформулированы следующие инициативы, учитывающие необходимость руководящих принципов для развития навыков в области квантовых вычислений и хорошо подготовленной рабочей силы для этой известной отрасли.

1) Определить степень детализации, глубины и сложности квантовых тем, которые должны преподаваться на каждом образовательном уровне.
2) Сопоставить результаты этого исследования с новейшими версиями стандарта SWEBOK.

References

[1]. S. Aralikatti, “Quantum Computing: Challenges and Opportunities,” in 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2021, pp. 1–4. doi: 10.1109/ICECCT52121.2021.9616647.

[2]. National Academies of Sciences, Engineering, and and Medicine, Quantum Computing: Progress and Prospects. Washington, DC: The National Academies Press, 2019.

[3]. M. Piattini, G. Peterssen, and R. Pérez-Castillo, “Quantum Computing: A New Software Engineering Golden Age,” SIGSOFT Softw. Eng. Notes, vol. 45, no. 3, pp. 12–14, Jul. 2020, doi: 10.1145/3402127.3402131.

[4]. B. Lee, C. Y. Liu, and M. L. J. Apuzzo, “Quantum Computing: A Prime Modality in Neurosurgery’s Future,” World Neurosurg, vol. 78, no. 5, pp. 404–408, 2012, doi: https://doi.org/10.1016/j.wneu.2012.07.013.

[5]. S. T. Marella and H. S. K. Parisa, “Introduction to Quantum Computing,” in Quantum Computing and Communications, Y. Zhao, Ed., Rijeka: IntechOpen, 2020, p. Ch. 5. doi: 10.5772/intechopen.94103.

[6]. J. Singh and M. Singh, “Evolution in Quantum Computing,” in 2016 International Conference System Modeling & Advancement in Research Trends (SMART), 2016, pp. 267–270. doi: 10.1109/SYSMART.2016.7894533.

[7]. Z. MENG, “Review of Quantum Computing,” in 2020 13th International Conference on Intelligent Computation Technology and Automation (ICICTA), 2020, pp. 210–213. doi: 10.1109/ICICTA51737.2020.00051.

[8]. Draup, “Quantum Computing Talent Ecosystem Analysis,” Nov. 2020.

[9]. E. Rieffel and W. Polak, QUANTUM COMPUTING A Gentle Introduction. Cambridge, Massachusetts: The MIT Press, 2011.
[10]. N. S. Yanofsky and M. A. Mannucci, Quantum computing for computer scientists. New York, NY: Cambridge University Press, 2008.
[11]. T. G. Wong, Introduction to Classical and Quantum Computing. Omaha, Nebraska: Rooted Grove, 2022.
[12]. E. Rieffel and W. Polak, "An Introduction to Quantum Computing for Non-Physicists," ACM Comput. Surv., vol. 32, no. 3, pp. 300–335, Sep. 2000, doi: 10.1145/367701.367709.
[13]. B. Bungum and S. Selstø, "What do quantum computing students need to know about quantum physics?," Eur J Phys, vol. 43, no. 5, p. 055706, 2022, doi: 10.1088/1361-6404/ac7e8a.
[14]. N. Bohr, "I. On the constitution of atoms and molecules," The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 26, no. 151, pp. 1–25, Jul. 1913, doi: 10.1080/14786441308634955.
[15]. W. Heisenberg, "On the quantum theory of line structure and anomalous Zeeman effect," Journal of Physics, vol. 8, Dec. 1922.
[16]. E. Schrödinger, "Quantization as an eigenvalue problem (Part I)," Annalen der Physics, vol. 384, no. 4, pp. 361–376, 1926.
[17]. P. Dirac, "The quantum theory of the electron," Proceedings of the Royal Society Series A, vol. 117, no. 778, p. 610-624, Feb. 1928.
[18]. E. E. Edwards, "Key Concepts for Future QIS Learners", Illinois Quantum Information Science and Technology Center.
[19]. G. Arun and V. Mishra, "A review on quantum computing and communication," in 2014 2nd International Conference on Emerging Technology Trends in Electronics, Communication and Networking, 2014, pp. 1–5. doi: 10.1109/ET2ECN.2014.7044953.
[20]. V. Verma, "A Study on Quantum Cryptography and Its Need," in Quantum Computing: A Shift from Bits to Qubits, R. Pandey, N. Srivastava, N. K. Singh, and K. Tyagi, Eds., Singapore: Springer Nature Singapore, 2023, pp. 407–435. doi: 10.1007/978-981-19-9530-9_21.
[21]. J. Preskill, "Simulating quantum field theory with a quantum computer," in The 36th Annual International Symposium on Lattice Field Theory -LATTICE2018, A. Bazavov, A. X. El-Khadra, S. Gottlieb, R. Lewis, H.-W. Lin, K.-F. Liu, Y. Meurice, J. Osbor, and A. Shindler, Eds., East Lansing, Michigan, USA.: Proceedings of Science, Nov. 2018, pp. 1–22.
[22]. V. Armaos, D. A. Badounas, P. Deligiannis, and K. Lianos, "Computational chemistry on quantum computers," Applied Physics A, vol. 126, no. 8, p. 625, 2020, doi: 10.1007/s00339-020-03755-4.
[23]. N. Kiefl and G. Hagel, "Software Engineering Education of Classical Computing vs. Quantum Computing: A Competency-Centric Approach," in Proceedings of the 4th European Conference on Software Engineering Education, in ECSEE '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 27–31. doi: 10.1145/3396802.3396816.
[24]. C. Cartiere, Quantum Software Engineering: Introducing Formal Methods into Quantum Computing. 2016.
[25]. M. Paltenghi and M. Pradel, "Bugs in Quantum Computing Platforms: An Empirical Study," Proc. ACM Program. Lang., vol. 6, no. OOPSLA1, Apr. 2022, doi: 10.1145/3527330.
[26]. R. Shaydulin, C. Thomas, and P. Rodeghero, "Making Quantum Computing Open: Lessons from Open Source Projects," in Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops, in ICSEW'20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 451–455. doi: 10.1145/3387940.3391471.
[27]. "QP4SE 2022: Proceedings of the 1st International Workshop on Quantum Programming for Software Engineering," New York, NY, USA: Association for Computing Machinery, 2022.
[28]. J. Zhao, Quantum Software Engineering: Landscapes and Horizons. 2020.
[29]. M. A. Akbar, A. A. Khan, S. Mahmood, and S. Rafi, "Quantum Software Engineering: A New Genre of Computing," Nov. 2022, Accessed: Mar. 18, 2023. [Online]. Available: https://arxiv.org/abs/2211.13990
[30]. A. Sarkar, "Automated Quantum Software Engineering: why? what? how?", Dec. 2022, doi: 10.48550/arxiv.2212.00619.
[31]. M. Piattini and J. M. Murillo, "Quantum Software Engineering Landscape and Challenges," in Quantum Software Engineering, M. A. Serrano, R. Pérez-Castillo, and M. Piattini, Eds., Cham: Springer International Publishing, 2022, pp. 25–38. doi: 10.1007/978-3-031-05324-5_2.
[32]. M. De Stefano, F. Pecorelli, D. Di Nucci, F. Palomba, and A. De Lucia, "Software Engineering for Quantum Programming: How Far Are We?", Mar. 2022, doi: 10.48550/arxiv.2203.16969.

[33]. A. Cobb, J.-G. Schneider, and K. Lee, “Towards Higher-Level Abstractions for Quantum Computing,” in Proceedings of the 2022 Australasian Computer Science Week, in ACSW ’22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 115–124. doi: 10.1145/3511616.3513106.

[34]. R. Juárez-Ramírez, S. Jiménez, C. X. Navarro, C. Guerra-García, H. G. Perez-Gonzalez, C. Fernández-y-Fernández, J. Ortiz-Hernández, and K. Cancino, “Skills required for Quantum Computing: A comprehensive review of recent Studies”, Programming and Computer Software, vol. 50, no. 8, 2024.

[35]. R. Abreu, S. Ali, T. Yue, M. Felderer, and I. Exman, “Quantum Software: Model-Driven or Search-Driven? A Q-SE 2021 Workshop Report,” SIGSOFT Softw. Eng. Notes, vol. 46, no. 4, pp. 23–25, Oct. 2021, doi: 10.1145/3485952.3485958.

[36]. C. Hughes, D. Finke, D.-A. German, C. Merzbacher, P. M. Vora, and H. J. Lewandowski, “Assessing the Needs of the Quantum Industry,” Aug. 2021, doi: 10.48550/arxiv.2109.03601.

[37]. A. Purohit, M. Kaur, Z. C. Seskir, M. T. Posner, and A. Venegas-Gomez, “Building a quantum-ready ecosystem,” IET Quantum Communication, vol. 5, no. 1, pp. 1–18, Mar. 2024, doi: https://doi.org/10.1049/qtc2.12072.

[38]. C. Singh, A. T. Asfaw, and J. Levy, “Preparing students to be leaders of the quantum information revolution,” Phys Today, 2021.

[39]. A. German, M. Pias, and Q. Xiang, “On the Design and Implementation of a Quantum Architectures Knowledge Unit for a CS Curriculum,” in Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1, in SIGCSE 2023. New York, NY, USA: Association for Computing Machinery, 2023, pp. 1150–1156. doi: 10.1145/3545945.3569845.

[40]. P. Bourque and R. E. Fairley, “Guide to the Software Engineering Body of Knowledge, Version 3.0,” 2014.

[41]. L. Gatti and R. Sotelo, “Quantum Computing for Undergraduate Engineering Students: Report of an Experience,” in 2021 IEEE International Conference on Quantum Computing and Engineering (QCE), 2021, pp. 397–401. doi: 10.1109/QCE52317.2021.00060.

[42]. A. Miranskyy, M. Khan, J. P. L. Faye, and U. C. Mendes, “Quantum computing for software engineering: prospects,” in Proceedings of the 1st International Workshop on Quantum Programming for Software Engineering, in QP4SE 2022. New York, NY, USA: Association for Computing Machinery, 2022, pp. 22–25. doi: 10.1145/3549036.3562060.

[43]. N. Abbas, J. Andersson, and D. Weyns, “ASPLe: A methodology to develop self-adaptive software systems with systematic reuse,” Journal of Systems and Software, vol. 167, p. 110626, 2020, doi: https://doi.org/10.1016/j.jss.2020.110626.

[44]. A. Filieri et al., “Control Strategies for Self-Adaptive Software Systems,” ACM Trans. Auton. Adapt. Syst., vol. 11, no. 4, Feb. 2017, doi: 10.1145/3024188.

[45]. H. Alsolai and M. Roper, “A systematic literature review of machine learning techniques for software maintainability prediction,” Inf Softw Technol, vol. 119, p. 106214, 2020, doi: https://doi.org/10.1016/j.infsof.2019.106214.

[46]. Y. Liu, S. Arunachalam, and K. Temme, “A rigorous and robust quantum speed-up in supervised machine learning,” Nat Phys, vol. 17, no. 9, pp. 1013–1017, Jul. 2021, doi: 10.1038/s41567-021-01287-z.

[47]. B. S. Ahmed, “Test case minimization approach using fault detection and combinatorial optimization techniques for configuration-aware structural testing,” Engineering Science and Technology, an International Journal, vol. 19, no. 2, pp. 737–753, 2016, doi: https://doi.org/10.1016/j.jestch.2015.11.006.

[48]. C. Gambella and A. Simonetto, “Multiblock ADMM Heuristics for Mixed-Binary Optimization on Classical and Quantum Computers,” IEEE Transactions on Quantum Engineering, vol. 1, pp. 1–22, 2020, doi: 10.1109/TQE.2020.3033139.

[49]. L. Braine, D. J. Egger, J. Glick, and S. Woerner, “Quantum Algorithms for Mixed Binary Optimization Applied to Transaction Settlement,” IEEE Transactions on Quantum Engineering, vol. 2, pp. 1–8, 2021, doi: 10.1109/TQE.2021.3063635.

[50]. F. Horváth, T. Gergely, Á. Beszédes, D. Tengeri, G. Balogh, and T. Gyimóthy, “Code coverage differences of Java bytecode and source code instrumentation tools,” Software Quality Journal, vol. 27, no. 1, pp. 79–123, 2019, doi: 10.1007/s11219-017-9389-z.

[51]. A. Miranskyy, A. Hamou-Lhadj, E. Cialini, and A. Larsson, “Operational-Log Analysis for Big Data Systems: Challenges and Solutions,” IEEE Softw, vol. 33, no. 2, pp. 52–59, 2016, doi: 10.1109/MS.2016.33.

[52]. J. Sevilla and C. Riedel, Forecasting timelines of quantum computing. 2020.

[53]. T. Iqbal, P. Elahidoost, and L. Lúcio, "A Bird's Eye View on Requirements Engineering and Machine Learning," in 2018 25th Asia-Pacific Software Engineering Conference (APSEC), 2018, pp. 11–20. doi: 10.1109/APSEC.2018.00015.

[54]. T. L. Frantz, "A Conversation with Terrill Frantz, Associate Professor of eBusiness and Cybersecurity, Harrisburg University," Jun. 04, 2021.

[55]. T. Plunkett, T. L. Frantz, H. Khatri, P. Rajendran, and S. Midha, "A Survey of Educational Efforts to Accelerate a Growing Quantum Workforce," in 2020 IEEE International Conference on Quantum Computing and Engineering (QCE), 2020, pp. 330–336. doi: 10.1109/QCE49297.2020.00048.

[56]. Y. Sedelmaier and D. Landes, "A Research Agenda for Identifying and Developing Required Competencies in Software Engineering," International Journal of Engineering Pedagogy, vol. 3, pp. 30–35, Apr. 2013, doi: 10.3991/ijep.v3i2.2448.

[57]. B. Bloom, M. Engelhart, E. Furst, W. Hill, and D. Krathwohl, Taxonomy of educational objectives: the classification of educational goals. Handbook I: cognitive domain. New York: McKay and Longman, 1956.

[58]. L. Anderson and D. Krathwohl, A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. 2001.

[59]. K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic Mapping Studies in Software Engineering," pp. 1–10, 2008, doi: 10.14236/ewic/EASE2008.8.

[60]. B. Barn, S. Barat, and T. Clark, "Conducting Systematic Literature Reviews and Systematic Mapping Studies," in Proceedings of the 10th Innovations in Software Engineering Conference, in ISEC '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 212–213. doi: 10.1145/3021460.3021489.

[61]. K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," Inf Softw Technol, vol. 64, pp. 1–18, 2015, doi: https://doi.org/10.1016/j.infsof.2015.03.007.

[62]. B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Technical report, EBSE Technical Report EBSE-2007-01, 2007. [Online]. Available: https://www.cs.auckland.ac.nz/~norsaremah/2007%20Guidelines%20for%20performing%20SLR%20in%20SE%20v2.3.pdf

[63]. B. A. Kitchenham, "Systematic reviews," in 10th International Symposium on Software Metrics, 2004. Proceedings., 2004, pp. xii–xii. doi: 10.1109/METRIC.2004.1357885.

[64]. J. Hannay, D. Sjøberg, and T. Dybå, "A Systematic Review of Theory Use in Software Engineering Experiments," Software Engineering, IEEE Transactions on, vol. 33, pp. 87–107, Mar. 2007, doi: 10.1109/TSE.2007.12.

[65]. G. Tebes, D. Peppino, P. Becker, and L. Olsina, Enhancing the Process Specification for Systematic Literature Reviews. 2019. doi: 10.13140/RG.2.2.14262.96321/1.

[66]. P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," Journal of Systems and Software, vol. 80, no. 4, pp. 571–583, 2007, doi: https://doi.org/10.1016/j.jss.2006.07.009.

[67]. B. Kitchenham, E. Mendes, and G. Travassos, "Cross versus Within-Company Cost Estimation Studies: A Systematic Review," Software Engineering, IEEE Transactions on, vol. 33, pp. 316–329, Jun. 2007, doi: 10.1109/TSE.2007.1001.

[68]. M. R. W. Hiebl, "Sample Selection in Systematic Literature Reviews of Management Research," Organ Res Methods, vol. 26, no. 2, pp. 229–261, Jan. 2021, doi: 10.1177/1094428120986851.

[69]. V. Garousi, M. Felderer, and M. V Mäntylä, "Guidelines for including grey literature and conducting multivocal literature reviews in software engineering," Inf Softw Technol, vol. 106, pp. 101–121, 2019, doi: https://doi.org/10.1016/j.infsof.2018.09.006.

[70]. B. Weder, J. Barzen, F. Leymann, and D. Vietz, "Quantum Software Development Lifecycle," Jun. 2021, doi: 10.48550/arxiv.2106.09323.

[71]. A. A. Khan et al., "Software Architecture for Quantum Computing Systems -- A Systematic Review," Feb. 2022, Accessed: Mar. 18, 2023. [Online]. Available: http://arxiv.org/abs/2202.05505

[72]. M. A. Serrano, J. A. Cruz-Lemus, R. Perez-Castillo, and M. Piattini, "Quantum Software Components and Platforms: Overview and Quality Assessment," ACM Comput. Surv., vol. 55, no. 8, Dec. 2022, doi: 10.1145/3548679.

[73]. L. Westfall and A. Leider, "Teaching Quantum Computing: Volume 2," 2019, pp. 63–80. doi: 10.1007/978-3-030-02683-7_6.

[74]. R. Juárez-Ramírez et al., “A Taxonomic View of the Fundamental Concepts of Quantum Computing–A Software Engineering Perspective,” Programming and Computer Software, vol. 49, no. 8, pp. 682–704, 2023, doi: 10.1134/S0361768823080108.

[75]. M. Mykhailova and K. M. Svore, “Teaching Quantum Computing through a Practical Software-Driven Approach: Experience Report,” in Proceedings of the 51st ACM Technical Symposium on Computer Science Education, in SIGCSE ’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1019–1025. doi: 10.1145/3328778.3366952.

[76]. M. A. Cusumano, “The business of quantum computing,” Commun. ACM, vol. 61, no. 10, pp. 20–22, Sep. 2018, doi: 10.1145/3267352.

[77]. S. Seegerer, T. Michaeli, and R. Romeike, “Quantum Computing As a Topic in Computer Science Education,” in Proceedings of the 16th Workshop in Primary and Secondary Computing Education, in WiPSCE ’21. New York, NY, USA: Association for Computing Machinery, 2021. doi: 10.1145/3481312.3481348.

[78]. R. Abreu, S. Ali, and T. Yue, “First International Workshop on Quantum Software Engineering (Q-SE 2020),” SIGSOFT Softw. Eng. Notes, vol. 46, no. 2, pp. 30–32, Mar. 2021, doi: 10.1145/3448992.3449000.

[79]. J. D. Weisz, M. Ashoori, and Z. Ashktorab, “Entanglion: A Board Game for Teaching the Principles of Quantum Computing,” in Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play, in CHI PLAY ’18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 523–534. doi: 10.1145/3242671.3242696.

[80]. M. De Vincentiis, F. Cassano, A. Pagano, and A. Piccinno, “QAI4ASE: Quantum artificial intelligence for automotive software engineering,” in Proceedings of the 1st International Workshop on Quantum Programming for Software Engineering, in QP4SE 2022. New York, NY, USA: Association for Computing Machinery, 2022, pp. 19–21. doi: 10.1145/3549036.3562059.

[81]. P. P. Angara, U. Stege, H. A. Müller, and M. Bozzo-Rey, “Hybrid quantum-classical problem solving in the NISQ era,” in Proceedings of the 30th Annual International Conference on Computer Science and Software Engineering, in CASCON ’20. USA: IBM Corp., 2020, pp. 247–252.

[82]. Z. Ashktorab, J. D. Weisz, and M. Ashoori, “Thinking Too Classically: Research Topics in Human-Quantum Computer Interaction,” in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, in CHI ’19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–12. doi: 10.1145/3290605.3300486.

[83]. B. C. Sanders, “Building a quantum computer (invited),” in Proceedings of the Workshop on System-Level Interconnect: Problems and Pathfinding Workshop, in SLIP ’20. New York, NY, USA: Association for Computing Machinery, 2020. doi: 10.1145/3414622.3431913.

[84]. D. Franklin et al., “Exploring Quantum Reversibility with Young Learners,” in Proceedings of the 2020 ACM Conference on International Computing Education Research, in ICER ’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 147–157. doi: 10.1145/3372782.3406255.

[85]. T. Gabor et al., “The Holy Grail of Quantum Artificial Intelligence: Major Challenges in Accelerating the Machine Learning Pipeline,” in Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops, in ICSEW’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 456–461. doi: 10.1145/3387940.3391469.

[86]. H. Neeman et al., “Cyberinfrastructure Facilitation Skills Training via the Virtual Residency Program,” in Practice and Experience in Advanced Research Computing, in PEARC ’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 421–428. doi: 10.1145/3311790.3396629.

[87]. C. de O. Melo and T. C. de Sousa, “Reflections on cyberethics education for millennial software engineers,” in Proceedings of the 1st International Workshop on Software Engineering Curricula for Millennials, in SECM ’17. IEEE Press, 2017, pp. 40–46. doi: 10.1109/SECM.2017.10.

[88]. T. Liu, D. Gonzalez-Maldonado, D. B. Harlow, E. E. Edwards, and D. Franklin, “Qupcakery: A Puzzle Game that Introduces Quantum Gates to Young Learners,” in Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1, in SIGCSE 2023. New York, NY, USA: Association for Computing Machinery, 2023, pp. 1143–1149. doi: 10.1145/3545945.3569837.

[89]. O. Weiser, Y. M. Kalman, C. Kent, and G. Ravid, “65 competencies: which ones should your data analytics experts have?”, Commun. ACM, vol. 65, no. 3, pp. 58–66, Feb. 2022, doi: 10.1145/3467018.

[90]. W. Auccahuasi, G. B. Santiago, E. O. Núñez, and F. Sernaque, “Interactive online tool as an instrument for learning mathematics through programming techniques, aimed at high school students,” in Proceedings

of the 6th International Conference on Information Technology: IoT and Smart City, in ICIT '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 70–76. doi:10.1145/3301551.3301580.

[91]. N. Dneprovskaya, S.-B. Chris Kang, and I. Shevtsova, "Evolution of the Competencies to Embrace Digital Technology for Sustainable Development," in Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, in CHI EA '22. New York, NY, USA: Association for Computing Machinery, 2022. doi: 10.1145/3491101.3519730.

[92]. C. Teuscher, "A golden age for computing frontiers, a dark age for computing education?", in Proceedings of the 18th ACM International Conference on Computing Frontiers, in CF '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 140–143. doi: 10.1145/3457388.3458673.

[93]. W. Mauerer, S. Klessinger, and S. Scherzinger, "Beyond the badge: reproducibility engineering as a lifetime skill," in Proceedings of the 4th International Workshop on Software Engineering Education for the Next Generation, in SEENG '22. New York, NY, USA: Association for Computing Machinery, 2023, pp. 1–4. doi: 10.1145/3528231.3528359.

[94]. I. Arawjo, "To Write Code: The Cultural Fabrication of Programming Notation and Practice," in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, in CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–15. doi: 10.1145/3313831.3376731.

[95]. M. Pias, B. Becker, Q. Xiang, M. Zahran, and M. Anderson, "Should Quantum Processor Design be Considered a Topic in Computer Architecture Education?", in Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2, in SIGCSE 2022. New York, NY, USA: Association for Computing Machinery, 2022, p. 1184. doi: 10.1145/3478432.3499201.

[96]. Y. Huang and M. Martonosi, "Statistical assertions for validating patterns and finding bugs in quantum programs," in Proceedings of the 46th International Symposium on Computer Architecture, in ISCA '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 541–553. doi: 10.1145/3307650.3322213.

[97]. J. Fitzjohn, A. Winckles, G. Wilson, and D. Vicinanza, "A Software Development Kit and Translation Layer for Executing Intel 8080 Assembler on a Quantum Computer (August 2022)," IEEE Transactions on Quantum Engineering, vol. 3, pp. 1–12, 2022, doi: 10.1109/TQE.2022.3204653.

[98]. S. Beniwal, "Need and Challenges in Quantum Computing in Fog Environment," in 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), 2023, pp. 175–180.

[99]. K. Prateek and S. Maity, "Quantum Programming on Azure Quantum—An Open Source Tool for Quantum Developers," in Quantum Computing: A Shift from Bits to Qubits, R. Pandey, N. Srivastava, N. K. Singh, and K. Tyagi, Eds., Singapore: Springer Nature Singapore, 2023, pp. 283–309. doi: 10.1007/978-981-19-9530-9_16.

[100]. S. Smuts and H. Smuts, "Society 5.0 and the future of work skills for software engineers and developers," in Proceedings of the Society 5.0 Conference 2022 - Integrating Digital World and Real World to Resolve Challenges in Business and Society, K. Hinkelmann and A. Gerber, Eds., in EPiC Series in Computing, vol. 84. EasyChair, 2022, pp. 169–182. doi: 10.29007/9kzd.

[101]. A. Bayerstadler et al., "Industry quantum computing applications," EPJ Quantum Technol, vol. 8, no. 1, p. 25, 2021, doi: 10.1140/epjqt/s40507-021-00114-x.

[102]. R. B. Shapiro and M. Tissenbaum, "New Programming Paradigms," The Cambridge Handbook of Computing Education Research, 2019, [Online]. Available: https://api.semanticscholar.org/CorpusID:86591035

[103]. J. Hooyberghs, Introducing Microsoft Quantum Computing for Developers: Using the Quantum Development Kit and Q#. 2021. doi: 10.1007/978-1-4842-7246-6.

[104]. A. Razak et al., "Chapter Seven - Reigniting the power of artificial intelligence in education sector for the educators and students competence", in Artificial Intelligence and Machine Learning in Smart City Planning, V. Basetti, C. K. Shiva, M. R. Ungarala, and S. S. Rangarajan, Eds., Elsevier, 2023, pp. 103–116. doi: https://doi.org/10.1016/B978-0-323-99503-0.00009-0.

[105]. F. Amato et al., "QuantuMoonLight: A low-code platform to experiment with quantum machine learning", SoftwareX, vol. 22, p. 101399, 2023, doi: https://doi.org/10.1016/j.softx.2023.101399.

[106]. R. Kurt, "Industry 4.0 in Terms of Industrial Relations and Its Impacts on Labour Life", Procedia Comput Sci, vol. 158, pp. 590–601, 2019, doi: https://doi.org/10.1016/j.procs.2019.09.093.

[107]. R. Verdecchia, P. Lago, and C. de Vries, "The future of sustainable digital infrastructures: A landscape of solutions, adoption factors, impediments, open problems, and scenarios," Sustainable Computing: Informatics and Systems, vol. 35, p. 100767, 2022, doi: https://doi.org/10.1016/j.suscom.2022.100767.

[108]. N. O. E. Olsson, E. Arica, R. Woods, and J. A. Madrid, “Industry 4.0 in a project context: Introducing 3D printing in construction projects,” Project Leadership and Society, vol. 2, p. 100033, 2021, doi: https://doi.org/10.1016/j.plas.2021.100033.
[109]. G. Morra, D. A. Yuen, H. M. Tufo, and M. G. Knepley, “Fresh Outlook on Numerical Methods for Geodynamics. Part 2: Big Data, HPC, Education” in Encyclopedia of Geology (Second Edition), D. Alderton and S. A. Elias, Eds., Oxford: Academic Press, 2021, pp. 841–855. doi: https://doi.org/10.1016/B978-0-08-102908-4.00111-9.
[110]. F. Greinert and R. Müller, Competence Framework for Quantum Technologies, Version 2.0. Quantum Flagship, 2023.
[111]. M. A. Nielsen, “Quantum information theory.” 2000.
[112]. M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, 10th edition. Cambridge, United Kingdom: Cambridge University Press, 2016.
[113]. C. E. Shannon and W. Weaver, The Mathematical Theory of Communication, 16th edition. The University of Illinois Press, 1971.
[114]. E. Sakai, “On the principles of quantum mechanics,” May 2004, doi: 10.48550/arxiv.quant-ph/0405069.
[115]. A. Karel Velan, “Basic Principles of Quantum Mechanics,” in The Multi-Universe Cosmos: The First Complete Story of the Origin of the Universe, A. Karel Velan, Ed., Boston, MA: Springer US, 1992, pp. 21–33. doi: 10.1007/978-1-4684-6030-8_3.
[116]. J. Bub, “Quantum Mechanics as a Principle Theory,” Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, vol. 31, no. 1, pp. 75–94, 2000, doi: https://doi.org/10.1016/S1355-2198(99)00032-5.
[117]. J. J. Sakurai, Modern Quantum Mechanics, 1st edition. Reading, MA: Addison-Wesley, 1994.
[118]. J. Singh and K. S. Bhangu, “Contemporary Quantum Computing Use Cases: Taxonomy, Review and Challenges,” Archives of Computational Methods in Engineering, vol. 30, no. 1, pp. 615–638, 2023, doi: 10.1007/s11831-022-09809-5.
[119]. R. Feynman, R. Leighton, M. Sands, and B. Lindsay, “The Feynman Lectures on Physics, Vol. 3: Quantum Mechanics,” Phys Today, vol. 19, no. 11, 1966.
[120]. S. Gasiorowicz, Quantum Physics, 2nd. edition. New York: Wiley, 1995.
[121]. B. Bungum and S. Selstø, “What do quantum computing students need to know about quantum physics?”, Eur J Phys, vol. 43, no. 5, p. 055706, Sep. 2022, doi: 10.1088/1361-6404/ac7e8a.
[122]. E. Murina, “Math and Physics Tools for Quality Quantum Programming,” in Quality of Information and Communications Technology, M. Shepperd, F. Brito e Abreu, A. Rodrigues da Silva, and R. Pérez-Castillo, Eds., Cham: Springer International Publishing, 2020, pp. 263–273.
[123]. P. Heron and L. McNeil, “Phys21: Preparing Physics Students for 21st-Century Careers. A report by the Joint Task Force on Undergraduate Physics Programs,” Dec. 2016.
[124]. P. J. Denning, “Computing is a Natural Science,” Commun. ACM, vol. 50, no. 7, pp. 13–18, Jul. 2007, doi: 10.1145/1272516.1272529.
[125]. D. E. Knuth, “Computer Science and Its Relation to Mathematics,” The American Mathematical Monthly, vol. 81, no. 4, pp. 323–343, 1974, doi: 10.2307/2318994.
[126]. P. J. Denning et al., “Computing as a discipline,” Computer (Long Beach Calif), vol. 22, no. 2, pp. 63–70, 1989, doi: 10.1109/2.19833.
[127]. A. J. DuBrin, Essentials of Management, 9th ed. Mason, OH: South-Western Cengage Learning, 2012.
[128]. M. F. J. Fox, B. M. Zwickl, and H. J. Lewandowski, “Preparing for the quantum revolution -- what is the role of higher education?”, Jun. 2020, doi: 10.1103/PhysRevPhysEducRes.16.020131.
[129]. M. M. Robles, “Executive Perceptions of the Top 10 Soft Skills Needed in Today’s Workplace,” Business Communication Quarterly, vol. 75, no. 4, pp. 453–465, Oct. 2012, doi: 10.1177/1080569912460400.
[130]. C. D. Aiello et al., “Achieving a quantum smart workforce,” Quantum Sci Technol, vol. 6, no. 3, p. 030501, Jul. 2021, doi: 10.1088/2058-9565/abfa64.
[131]. QURECA Ltd, “A Review of Global Quantum Education Initiatives,” Apr. 2022.
[132]. M. N. Amin, R. P. Uhlig, P. P. Dey, B. Sinha, and S. Jawad, “The Needs and Challenges of Workforce Development in Quantum Computing,” American Society for Engineering Education, 2019.

Информация об авторах / Information about authors

Рейес ХУАРЕС-РАМИРЕС получил степень PhD по информатике в Автономном университете Нижней Калифорнии (UABC) в 2008 году по специальности «Разработка

программного обеспечения». В настоящее время занимает должность полного профессора в UABC, кампус Тихуаны, Нижняя Калифорния (Мексика). Президент Мексиканского общества разработки программного обеспечения (REDMIS, https://conisoft.org/redmis/). Член Национальной системы исследователей Мексики (SNI), уровень 2, руководит несколькими исследовательскими проектами, как академическими, так и в сотрудничестве с промышленностью. Сфера научных интересов: программная инженерия (неопределенность в гибких методологиях, улучшение качества в Scrum), человеко-машинное взаимодействие (проектирование, ориентированное на пользователя, адаптивные пользовательские интерфейсы), с недавних пор работает в области квантовых вычислений. Главный председательствующий на некоторых национальных и международных мероприятиях, таких как Международная конференция по исследованиям и инновациям в области разработки программного обеспечения (CONISOFT).

Reyes JUÁREZ-RAMÍREZ received a PhD from the Universidad Autónoma de Baja California (UABC) in 2008, in Computer Science with a specialty in Software Engineering. He is currently holding a Full Professor position in UABC, Tijuana campus, Baja California, Mexico. He is the President of the Mexican Network of Software Engineering (REDMIS, https://conisoft.org/redmis/). He is a member of the National System of Researchers of Mexico (SNI), Level 2, and leads several research projects, academic and in collaboration with industry. His research areas are Software Engineering (uncertainty in agile methodologies, quality improvement in Scrum), Human-Computer Interaction (user-centered design, adaptive user interfaces), and recently he has started working with Quantum Computing. He is the general chair of some national and international events, such as the International Conference in Software Engineering Research and Innovation (CONISOFT).

Саманта ХИМЕНЕС имеет многолетний опыт преподавания компьютерных технологий, программирования и информатики для студентов бакалавриата. Принимала участие в качестве лидера в различных проектах по разработке программного обеспечения как исследовательского, так и прикладного характера. Получила степень PhD в области информатики в Автономном университете Нижней Калифорнии, степень магистра в области инженерии и степень бакалавра в области инженерии компьютерных систем в Университете Колимы, участвует в Национальной системе исследователей Мексики. Её научными интересами являются взаимодействие человека с компьютером, диалоговые системы, аффективные вычисления и образовательные системы. Она представляла работы на нескольких конференциях и опубликовала более 20 статей в международных научных журналах.

Samantha JIMÉNEZ has several years of experience teaching computer software engineering, programming, and computer science for bachelor's-level students. She has participated in a leadership role in various software development projects both for industry and academia. She holds a PhD in Computer Sciences from the Autonomous University of Baja California, a master's in engineering and a bachelor's degree in computer systems Engineering from the University of Colima. She is part of the National System of Researchers of Mexico. Her research interests are human-computer interaction, dialogue systems, affective computing, and educational systems. She has presented her work in several conferences and published more than 20 articles in international scientific journals.

Кристиан Ксавьер НАВАРРО-КОТА получил степень магистра по информатике в Центре научных исследований и высшего образования в Энсенаде в 2002 году и степень PhD в Университете Кастилии-Ла-Манча (Испания) в 2016 году. Работает на факультете инженерии, архитектуры и дизайна в Автономном университете Нижней Калифорнии (Мексика) с 2001 года, в настоящее время является профессором и исследователем. Его научные интересы сосредоточены в областях мобильных вычислений, взаимодействия человека и компьютера, повсеместных вычислений и машинного обучения.

Christian Xavier NAVARRO-COTA – received a Computer Science master's degree from the Center of Scientific Research and Superior Education in Ensenada, in 2002, and a PhD degree from the University of Castilla–La Mancha, Spain, in 2016. He joined the Faculty of Engineering, Architecture and Design at the Universidad Autónoma de Baja California, Mexico, in 2001, where he is currently a Professor and Researcher. His research interests are focused on the fields of Mobile Computing, Human–Computer Interaction, Ubiquitous Computing, and M-learning.

Сезар ГУЕРРА-ГАРСИЯ – профессор кафедры вычислений Автономного университета Сан-Луис-Потоси (Мексика). Получил степень PhD в Университете Кастилии-Ла-Манча в 2011 году (Испания). Он получил свою первую степень в области компьютерных систем в 1996 году в Технологическом институте Сан-Луис-Потоси, Мехико. Имеет степень магистра информатики университета CICESE с 2000 года. Сфера научных интересов: разработка программного обеспечения, качество данных, разработка требований, разработка моделей, веб-инжиниринг и гражданская наука. Он имеет опыт преподавания более 20 лет в области разработки программного обеспечения и моделирования программного обеспечения.

César GUERRA-GARCÍA is currently Full Time Professor at Department of Computation, Autonomous University of San Luis Potosí, México. He obtained his PhD in University of Castilla-La Mancha in 2011, Spain. He did his First Degree in Computer Systems in 1996 in Technologic Institute of San Luis Potosí, México. He also obtained his Master of Computer Sciences at CICESE in 2000. His areas of interests are Software Engineering, Data Quality, Requirements Engineering, Model Driven Development, Web Engineering and Citizen Science. He has teaching experience of more than 20 years in the areas of Software Engineering and Software Modelling.

Эктор Херардо ПЕРЕС-ГОНСАЛЕС – профессор-исследователь Автономного университета Сан-Луис-Потоси (Мексика). Имеет степени магистра компьютерных наук (UNAM) и PhD по информатике (Университет Колорадо). Автор научно-исследовательских работ по автоматическому проектированию программного обеспечения и по квантовым вычислениям. Докладчик на конференциях (США, Бразилия, Канада, Англия, Испания, Португалия, Сингапур). Рецензент статей для международных журналов и конференций. Секретарь Мексиканского общества компьютерных наук, 2012-2014. Заведующий кафедрой информатики (Инженерная школа, УАСЛП). Лауреат всемирной стипендии Ли Кимче МакГрат 2014 года от Ассоциации научно-технических центров (ASTC). Член Национальной системы исследователей Мексики уровня I.

Hector Gerardo PEREZ-GONZALEZ is a research professor at the Autonomous University of San Luis Potosí. Master of Science in Computer Science (UNAM) and PhD in Computer Science (University of Colorado). Author of research papers on automatic software design and for quantum computers. Speaker at conferences (USA, Brazil, Canada, England, Spain, Portugal, and Singapore). Reviewer of papers for international journals and conferences. Secretary of the Mexican Society of Computer Science, 2012-2014. Head of the Computer Science department (Engineering school, UASLP). Recipient of the 2014 Lee Kimche McGrath Worldwide Fellowship from the Association of Science and Technology Centers (ASTC). Member of the National System of Researchers of Mexico Level I.

Карлос Альберто ФЕРНАНДЕС-И-ФЕРНАНДЕС имеет степень бакалавра по информатике Университета Веракруса, степень магистра информатики от Фонда Артуро Розенблуэта и степень PhD Университета Шеффилда в области разработки программного обеспечения. Штатный профессор и научный сотрудник Института вычислительной техники в Технологическом университете Миштека, где он также является директором. Член Мексиканской академии вычислительной техники (AMEXCOMP), академическая секция "Программная инженерия". Член Руководящего комитета Международного конгресса по исследованиям и инновациям в области разработки программного обеспечения (CONISOFT)

и Мексиканской сети разработки программного обеспечения (REDMIS). Член различных научных комитетов национальных и международных журналов и конференций. В области разработки программного обеспечения его основные направления включают как гибкие, так и традиционные методы разработки, визуальное моделирование и формальную спецификацию программного обеспечения.

Carlos Alberto FERNÁNDEZ-Y-FERNÁNDEZ holds a bachelor's degree in computer science from the Universidad Veracruzana, a master's degree in Computer Science from the Arturo Rosenblueth Foundation, and a PhD from the University of Sheffield in the field of Software Engineering. He is a full-time professor and researcher at the Institute of Computing in the Technological University of the Mixteca, where he also serves as the director. He is a member of the Mexican Academy of Computing (AMEXCOMP), in the Software Engineering academic section. He is a member of the Steering Committee of the International Congress on Research and Innovation in Software Engineering (CONISOFT) and the Mexican Network of Software Engineering (REDMIS). He is also a member of various scientific committees for national and international journals and conferences. Within the field of Software Engineering, his focus areas include both agile and traditional development methods, visual modeling, and formal software specification.

Хавьер ОРТИС-ЭРНАНДЕС имеет степень PhD по автоматизации и информатике, профессор кафедры программирования Национального центра исследований и технологических разработок (CENIDET) в Мексике. Научные интересы: Цифровая трансформация, разработка требований, моделирование бизнес-процессов.

Javier ORTIZ-HERNÁNDEZ – PhD in Automation and Computer Science, full professor at the Computer Science Department of the National Center of Research and Technological Development (CENIDET) in Mexico. Research interests: Digital transformation, requirements engineering, business process modeling.

Карина КАНСИНО – профессор по программной инженерии в Политехническом университете Тапачулы с большим опытом разработки приложений и управления моделями данных. Активный член Мексиканской сети разработки программного обеспечения (REDMIS) и Научного сообщества инженеров, работающих в прикладных областях. Она имеет признание Уровня I Национальной системы исследователей Мексики (SNI) области VIII, отличие в области Инженерного и технологического развития и звание почетного исследователя уровня VI Государственной системы исследователей штата Чьяпас. Ее основные направления исследований – моделирование баз знаний, применение описательных технологий для анализа данных и временны́е ряды.

Karina CANCINO is a full-time professor of Software Engineering at UPTapachula with extensive experience in application development and data model management. She is currently an active member of the Mexican Network of Software Engineering (REDMIS) and the Academic Body Engineering Applied to Organizations. She has the recognition Level I National System of Researchers of Mexico (SNI) Area VIII. Engineering and Technological Development and Honorary Researcher Level VI of the State System of Researchers of the State of Chiapas. Her main lines of research are knowledge base modeling, application of descriptive technologies for data analysis and time series.

DOI: 10.15514/ISPRAS-2024-36(6)-8

Specifying Cognitive Solutions in Complex Informally Structured Domains: Empirical Approaches

[1] *A. Jiménez-Galina, ORCID: 0000-0001-8713-0518 <al206560@alumnos.uacj.mx>*
[1] *K. Olmos-Sánchez, ORCID: 0000-0002-9145-6761 <kolmos@uacj.mx>*
[2] *A. Maldonado-Macias, ORCID: 0000-0002-4959-161X <amaldona@uacj.mx>*
[3] *J. Licona-Olmos, ORCID: 0000-0001-9487-6297 <jazmin@uaeh.edu.mx>*
[1] *P. Sánchez-Solís, ORCID: 0000-0001-5514-5061 <julia.sanchez@uacj.mx>*
[1] *Departamento de Ingeniería Eléctrica y Computación,*
Universidad Autónoma de Ciudad Juárez (UACJ), Chihuahua, Mexico
[2] *Departamento de Ingeniería Industrial y Manufactura,*
Universidad Autónoma de Ciudad Juárez (UACJ), Chihuahua, Mexico
[3] *Instituto de Ciencias Básicas e Ingeniería,*
Universidad Autónoma del Estado de Hidalgo (UAEH). Pachuca, Mexico

Abstract. The complexity of dealing with work-related stress as a Complex Informal Structured Domain (CISD) involves various social, technical, cultural, and scientific factors, which highlights the challenges posed by organizational decision-making and the need for cognitive solutions to improve understanding of such complex scenarios. The article discusses three empirical-theoretical approaches to conceptualizing and specifying cognitive solutions to real-world problems in CISD: A literature review examines the use of specific machine learning artificial algorithms to develop models for work stress prevention; the use of cognitive solutions as ontologies for explicit knowledge representation; and a systemic methodological framework that establishes a structured approach to conceptualization and specification. The exploration emphasizes the need for a methodological model that effectively supports these cognitive solutions, to improve organizational decision-making by leveraging systems thinking and knowledge management.

Keywords: cognitive solutions; knowledge management; systems thinking; complex domains.

For citation: Jiménez-Galina A., Olmos-Sánchez K., Maldonado-Macias A., Licona-Olmos J., Sánchez-Solís P. Specifying Cognitive Solutions in Complex Informally Structured Domains: Empirical Approaches. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 149-160. DOI: 10.15514/ISPRAS-2024-36(6)-8.

Acknowledgements. On behalf of all co-authors and collaborators, we express our most sincere gratitude and eternal admiration for Dr. Jazmín Georgina Licona Olmos. Her contribution was invaluable, and her legacy remains on these pages. She was a brilliant researcher, an exceptional mentor, and an inspiring person. The first author would also like to express her gratitude to the Consejo Nacional de Humanidades, Ciencias y Tecnologías (Conahcyt) for the fellowship support provided during my doctoral studies at the Programa de Doctorado en Ciencias de la Ingenieria Avanzada, Universidad Autónoma de Ciudad Juárez (UACJ).

Определение когнитивных решений в сложных неформально структурированных доменах: эмпирические подходы

[1] *А. Хименес-Галина, ORCID: 0000-0001-8713-0518 <al206560@alumnos.uacj.mx>*
[1] *К. Олмос-Санчес, ORCID: 0000-0002-9145-6761 <kolmos@uacj.mx>*
[2] *А. Мальдонадо-Мачиас, ORCID: 0000-0002-4959-161X <amaldona@uacj.mx>*
[3] *Ж. Ликона-Олмос, ORCID: 0000-0001-9487-6297 <jazmin@uaeh.edu.mx>*
[1] *П. Санчес-Солис, ORCID: 0000-0001-5514-5061 <julia.sanchez@uacj.mx>*

[1] *Факультет Электроинженерии и вычислений, Автономного университета Сьюдад-Хуареса (UACJ), Чиуауа, Мексика*
[2] *Факультет промышленности и производства, Автономного университета Сьюдад-Хуареса (UACJ), Чиуауа, Мексика*
[3] *Институт фундаментальных наук и инженерии, Автономный университет штата Идальго (UAEH), Пачука, Мексика*

Аннотация. Сложность борьбы со стрессом, связанным с работой, как сложной неформальной структурированной областью деятельности (CISD), включает в себя различные социальные, технические, культурные и научные факторы, что увеличивает организационные проблемы и подчеркивает необходимость когнитивных решений для улучшения понимания таких сложных сценариев. В статье рассматриваются три эмпирико-теоретических подхода к концептуализации и определению когнитивных решений реальных проблем в CISD: Авторами изучалась литература по использованию специфических алгоритмов машинного обучения нейронных сетей для разработки моделей профилактики стресса на работе; использованию когнитивных решений, в частности онтологий, для четкого представления знаний; также изучалась системная методологическая основа структурированного подхода к концептуализации и спецификации. Проведенное исследование показывает необходимость построения методологической модели, которая должна эффективно поддерживать эти когнитивные решения, для улучшения качества осуществляемых организационных мероприятий за счет использования системного мышления и управления знаниями.

Ключевые слова: когнитивные решения; управление знаниями; системное мышление; сложные домены.

Для цитирования: Хименес-Галина А., Олмос-Санчес К., Мальдонадо-Мачиас А., Ликона-Олмос Ж., Санчес-Солис П. Определение когнитивных решений в сложных неформально структурированных доменах: эмпирические подходы. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 149–160 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–8.

Благодарности. От имени всех соавторов и сотрудников мы выражаем нашу самую искреннюю благодарность и вечное восхищение доктору Жасмин Георгиане Ликона-Олмос. Ее вклад был неоценим, и ее наследие остается на этих страницах. Она была блестящим исследователем, исключительным наставником и вдохновляющим человеком. Первый автор также выражает свою благодарность Национальному совету по гуманитарным наукам и технологиям (Conahcyt) за стипендиальную поддержку, оказанную во время ее учебы в докторантуре по программе инженерных наук Автономного университета Сьюдад-Хуарес (UACJ).

1. Introduction

Complex domains are characterized by their dynamic nature, open boundaries, and ambiguous relationships, necessitating customized approaches due to their unique characteristics. These domains lack formal structures that feature non-linear and interdependent factors and are marked by uncertainty, diverse perspectives, and continuous change, with tacit knowledge being predominant. Understanding these domains is crucial for developing artificial intelligence (AI) based cognitive solutions, which formalize domain concepts and relationships, such as Decision Support Systems

and Predictive Analytics Tools. These solutions help organizations make informed decisions, drive innovation, and enhance performance by deeply comprehending complex situations.

Complex Informal Structured Domains (CISD) arise when external specialists, unfamiliar with the domain, are brought in to develop cognitive solutions. This introduces complexity but is essential for bringing fresh perspectives and innovative approaches. Challenges in CISD include knowledge elicitation, representation, reaching consensus among specialists, and designing effective solutions.

The complexity of CISD, characterized by informal knowledge, multiple perspectives, ambiguity, and uncertainty, complicates establishing clear relationships between concepts. For instance, work stress, a global issue with unpredictable behaviors like emotional exhaustion and symptoms such as anxiety and depression, affects organizational productivity and can also have positive impacts. Developing cognitive solutions involves navigating diverse data and viewpoints, making work stress particularly challenging (see Fig.1).

This article presents three approaches to developing cognitive solutions for work stress in CISD within the maquiladora industry on Mexico's northern border: (1) a literature review of the application of machine learning (ML) algorithms, to model work stress prevention; (2) using cognitive solutions with explicit knowledge representation, such as ontologies; and (3) the creating of a systemic methodological framework for conceptualizing and specifying cognitive solutions in CISD. As a result of this analysis, this paper introduces a methodological model that employs systemic thinking and knowledge management to support organizational decision-making. The paper is structured as follows: Section 2 covers CISD and work stress characterization, Section 3 reviews the approaches, and Section 4 discusses results and future work.

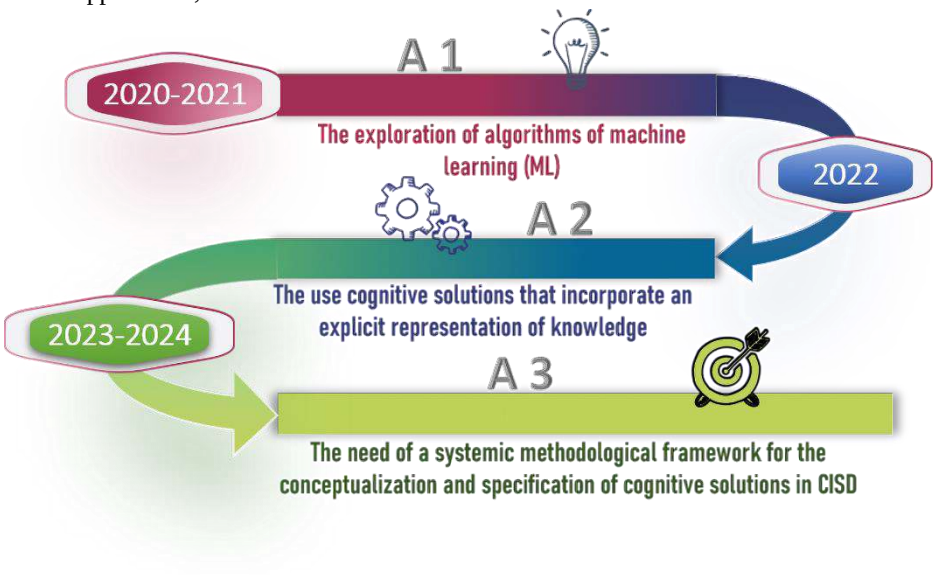


Fig. 1. Implemented approaches. Based on [1].

2. Background

2.1 Complex Informal Structured Domains

Complex Informal Structured Domains (CISD) arise when a team of domain experts and newcomers collaborate to create cognitive solutions tailored to the domain's needs. CISD involves multiple interacting elements, making reductionist approaches potentially ineffective. The key characteristics and components of CISD, illustrated in Fig. 2, include:

a) CISD is composed of an Application Domain (AD) and a Solution Domain (SD).
b) Domain Specialists (DS) with expertise in specific areas provide critical insights and guidance, although their perspectives may vary based on their roles.
c) The AD encompasses human, technological, cultural, and scientific factors, characterized by ambiguity, uncertainty, dynamism, and emergent properties.
d) Solution Providers, managed by a Cognitive Architect, specialize in conceptualizing, specifying, designing, and implementing cognitive solutions.
e) All participants form an ad hoc Collaborative Network.
f) Cognitive Solutions Specification includes both functional and non-functional requirements, derived from and synthesized between AD and SD.

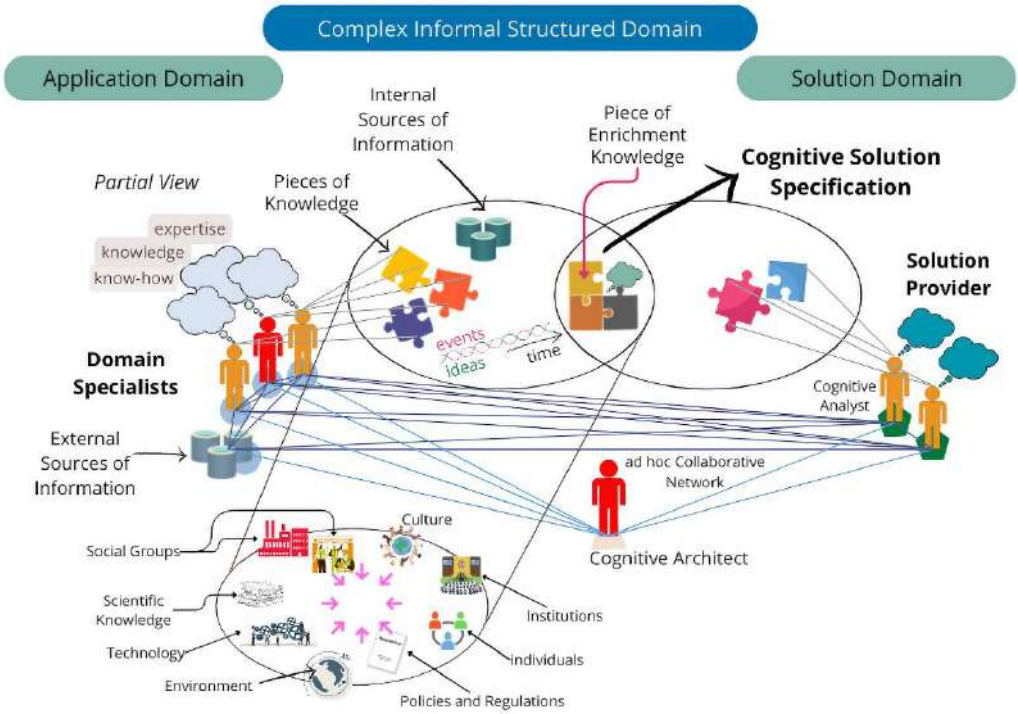


Fig. 1. CISD Characterization. Based on [2].

2.2 Work stress as CISD

Work stress impacts 75% of workers in Mexico, resulting in various disorders due to factors like work overload, harassment, insufficient support, and autonomy issues. The COVID-19 pandemic has worsened stress due to increased uncertainty. To address this, organizations such as the World Health Organization, National Human Rights Commission, and Secretaría del Trabajo y Previsión Social (México) recommend identifying psychosocial risks, implementing work-life balance policies, conducting risk assessments, and promoting mental well-being practices.

Effective intervention involves collaboration between the government, employers, and workers, with the government enforcing policies like NOM-035-STPS-2018, employers addressing risks, and workers engaging in management training. In the maquiladora industry, which encompasses human, technological, cultural, and scientific factors, work stress is influenced by various aspects including interpersonal dynamics, technological changes, and cultural and scientific research. Domain Specialists and Solution Providers work together in a collaborative network, addressing dynamic and emergent stress factors, which vary based on individual resilience and coping skills.

3. Approaches to addressing CISD

As was mentioned before, three approaches were analyzed:

1) a literature review of ML techniques,
2) cognitive solutions with explicit knowledge representation, and
3) a systemic methodological model for CISD.

3.1 Approach 1: Use of artificial intelligence techniques

The initial approach involved a literature review of cognitive solutions aimed at preventing work stress in the maquiladora industry through AI and ML. Databases like Science Direct and Taylor & Francis helped identify 21 relevant articles from a total of 61.

ML algorithms are increasingly applied to manage and prevent mental disorders, gaining traction in healthcare due to their ability to learn from data. Advances in data availability and computing have furthered ML research. ML is used in various domains, including gaming and mental health, to enhance data analysis and understanding of psychological conditions [3–5].

Despite these promising developments, the focus on work stress remains limited, and previous studies often lacked enough characterization of prior knowledge, which impacted the quality of the results. Recent research underscores the importance of explicit knowledge representation in cognitive system design. Peruvian regulations, for example, require ISO/IEC 12207 compliance for software requirements validation [6], and studies highlight the benefits of validation for quality and efficiency improvements [7]. Models like CMMI and ISO/IEC 15504 are essential for process improvement [8], with advancements in the Analytic Hierarchy Process (AHP) proposed to enhance prioritization accuracy [9].

ML faces challenges such as ambiguity, unclear definitions, incomplete data, and the need for domain-specific knowledge, necessitating the use of complementary techniques to address these issues effectively.

These insights emphasize the need for a more robust solution that integrates knowledge management and representation, particularly tacit knowledge, to effectively address work stress in complex domains like the maquiladora industry.

3.2 Approach 2: Cognitive solutions that have an explicit knowledge representation

For this approach, we applied the KMoS-RE (Knowledge Management on a Systematic Process for Requirements Engineering) approach to develop a knowledge representation of work stress in the maquiladora industry. KMoS-RE focuses on capturing and enhancing tacit knowledge in Informally Structured Domains (ISD) through cognitive techniques and a spiral knowledge evolution cycle [10]. It is based on the Knowledge Evolution Model for Requirements Engineering and Nonaka and Takeuchi's knowledge creation model [11] and includes domain modeling, system modeling, and requirements specification development [12]. During domain modeling, we created essential components such as the KDEL (Knowledge Domain Extended Lexicon), a conceptual model, and an OWL ontology, based on a review of standards and literature. Despite involving domain specialists for validation, several limitations were identified:

- KMoS-RE struggles with ambiguity and uncertainty, relying on consensus for defining concepts.
- A single model approach can delay understanding by lacking multiple perspectives.
- The knowledge representation is partial and does not fully capture the domain.

While KMoS-RE aids in explicit knowledge representation, a broader approach that includes diverse perspectives and cultural factors is necessary for effective cognitive solution development.

3.3 Approach 3: Systemic methodological framework for conceptualization and specification of cognitive solutions in CISD

In this approach, we use a systemic methodological framework to develop cognitive solutions in CISD, drawing on Morin's complexity theory [13] and Checkland's Soft Systems Methodology (SSM) [14-15]. Morin's theory focuses on understanding interconnected, nonlinear systems, while Checkland's SSM addresses complex, poorly defined problems through diverse perspectives and structured debate. Applying these approaches, we used a cause-effect problem tree to analyze CISD issues, identifying key challenges: neglect of tacit knowledge, poor communication, and incomplete domain observation [16].

The third approach focused on developing a systemic methodological model for conceptualizing and specifying cognitive solutions in Complex Informal Structured Domains (CISD). This model combines systems thinking with knowledge management principles to address domain complexities. Key results include:

- Model Development: The KMoS-SSA (Knowledge Management of Strategic options through Soft Systemic Analysis) framework merges SSM with the KMoS-RE process, allowing for flexible, evolving analysis and continuous solution adaptation based on new domain information [2].
- Systems Thinking: This approach assesses interrelationships and dynamics within CISD, crucial for tackling complex issues like work-related stress that involve multiple social, technical, and cultural factors.
- Enhanced Knowledge Elicitation: The framework improves the elicitation and representation of domain knowledge by integrating specialist perspectives and creating a shared language for addressing diverse needs.
- Requirements Specification: It establishes a clear process for defining both functional and non-functional requirements, ensuring solutions are desirable and feasible by deriving insights from interactions between application and solution domains.
- Positive Results: The model has effectively conceptualized cognitive solutions for work stress in the maquiladora industry, improving understanding through varied perspectives.

These findings support the proposed systemic approach for managing complexity in CISD. In summary, the proposed systemic approach provides a strong framework for developing and defining cognitive solutions within CISD. It enables more efficient handling of the intricate and evolving aspects of issues such as work stress.

4. Results

The KMoS-SSA model aims to conceptualize and specify cognitive solutions for CISD issues using systemic thinking. It enhances domain understanding and supports the development of effective, feasible solutions for organizational decision-making through its Knowledge Enrichment Cycle and methodological model [2].

4.1 Methodological Model for CISD: Knowledge Enrichment Cycle

The knowledge enrichment cycle model is an iterative framework that integrates systems thinking with real-world perspectives (Fig. 3). It combines the SECI model, the KMoS-RE knowledge evolution model, and systemic thinking principles. The SECI model involves four stages – socialization, externalization, combination, and internalization – to generate and interact with both tacit and explicit knowledge. The KMoS-RE model adapts these concepts for Requirements Elicitation. The cycle includes

a) Knowledge Elicitation, which gathers and organizes information;

b) Knowledge Enrichment, which validates and connects the information;

c) Model Generation, which creates artifacts representing the domain;

d) Model Discussion, which involves presenting models to specialists for feedback; and

e) Model Validation, where specialists assess and adjust the models.

This iterative process supports the continuous refinement of domain understanding and solution development.

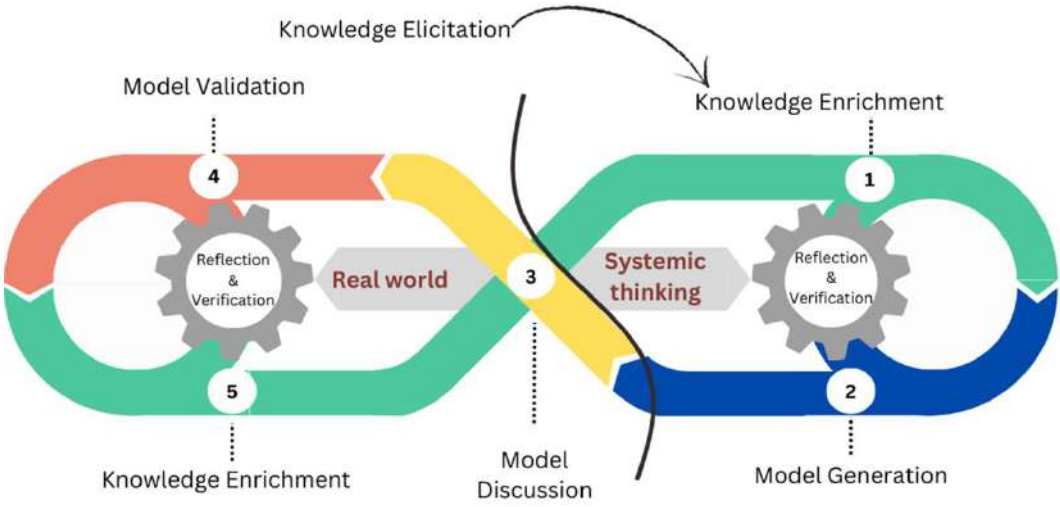


Fig. 3. Knowledge Enrichment Cycle. Based on [2].

4.2 Methodological Model for CISD: general activities, processes, and methods view

The CISD methodological model (Fig. 4) involves a set of activities and processes divided into real-world and systems-thinking tasks.

Real-world tasks cover ***elicitation*** and ***decision-making*** stages, while systems thinking tasks include ***structuring***, ***generation***, and ***enrichment*** stages [2]. Elicitation involves rigorous knowledge gathering, and decision-making includes validation and enrichment. Domain structuring formalizes the CISD, while model generation focuses on creating and validating artifacts. The enrichment stage involves thorough analysis and enhancement of knowledge.

5. Future work

We are currently working on the application of the proposed methodological model to the work stress phenomenon in the maquiladora industry, a real-world problem. We are also applying the methodology to generate solution alternatives in other areas such as academic accreditations and psychological therapies to provide evidence of its functionality.

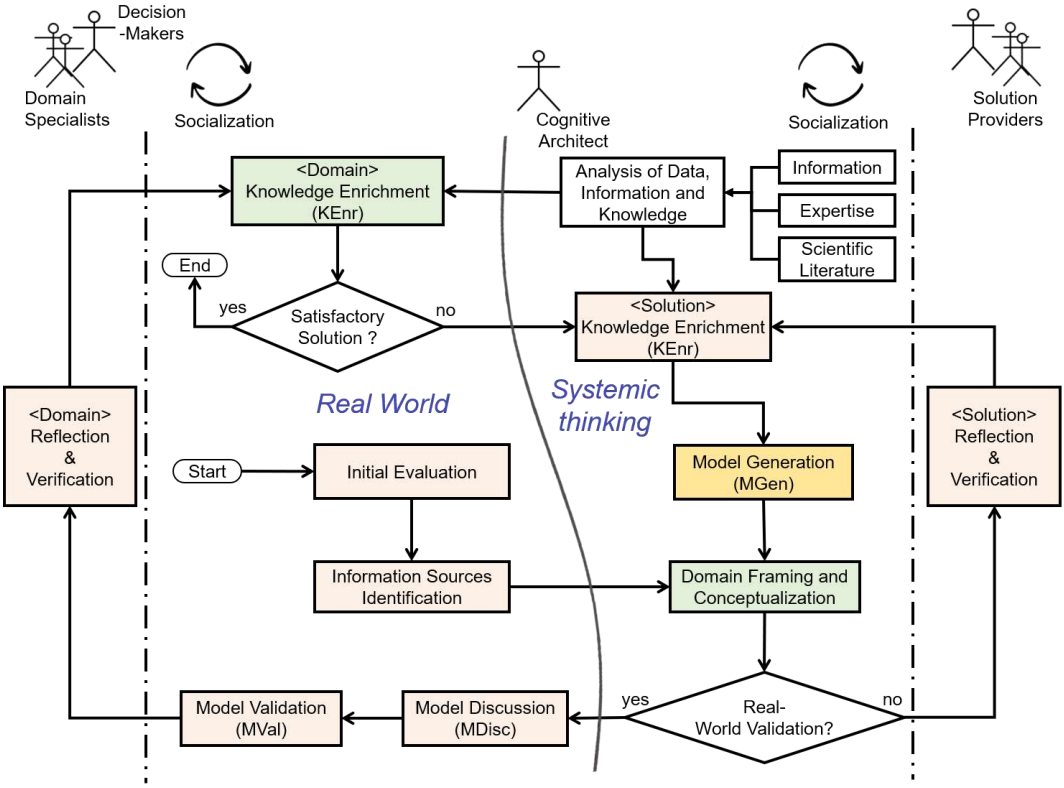


Fig. 4. Methodological model for CISD: general activities, processes, and methods view. Based on [2].

References

[1]. A. M. Jiménez-Galina, K. M. Olmos-Sánchez, J. G. Licona-Olmos, and A. A. Maldonado-Macias, "Necesidad de una Metodología de Representación del Conocimiento para Sistemas Complejos en Dominios de Estructura Informal: Artículo sobre enfoques empírico-teóricos" 11th Int. Conf. Softw. Eng. Res. Innov. (CONISOFT 2023)., 2023.

[2]. K. Olmos-Sánchez, J. Rodas-Osollo, A. Maldonado-Macías, A. Jiménez-Galina. Harmonization of knowledge representation: Integrating systems thinking ideas with appropriate domain representation strategies. Journal of Computación y Sistemas in New Trends in Artificial Intelligence and its Applications, 2024. (in Press).

[3]. N. Jaques, S. Taylor, E. Nosakhare, A. Sano, and R. Picard, "Multi-task Learning for Predicting Health, Stress, and Happiness", NIPS Work. Mach. Learn. Healthc., pp. 1–5, 2016.

[4]. S. Ware et al., "Machine Learning for Mental Health", J. Biomed. Inform., vol. 2, no. March, 2020.

[5]. K. Saha, B. Sugar, J. Torous, B. Abrahao, E. Kıcıman, and M. De Choudhury, "A social media study on the effects of psychiatric medication use", Proc. 13th Int. Conf. Web Soc. Media, ICWSM 2019, pp. 440–451, 2019, doi: 10.1609/icwsm.v13i01.3242.

[6]. Canchari, L. and Dávila, A., Requirements validation in the information systems software development: an empirical evaluation of its benefits for a Public Institution in Lima, in Trends and Applications in Software Engineering. CIMPS 2019. Advances in Intelligent Systems and Computing, Mejia, J., Munoz, M., and Rocha, A.A.C.-M.J., Eds., Guanajuato: Springer Nature Switzerland, 2020, pp. 23–35. https://doi.org/10.1007/978-3-030-33547-2_3.

[7]. Canchari, L., Angeleri, P. & Dávila, A. Requirements Validation in the Information System Software Development Lifecycle: A Software Quality in Use Evaluation. Program Comput Soft 49, 610–624 (2023). https://doi.org/10.1134/S0361768823080054.

[8]. Almeyda, S., Dávila, A. Process Improvement in Software Requirements Engineering: A Systematic Mapping Study. Program Comput Soft 48, 513–533 (2022). https://doi.org/10.1134/S0361768822080084
[9]. Somohano-Murrieta, J.C., Ocharán-Hernández, J.O., Sánchez-García, Á.J. et al. Improving the Analytic Hierarchy Process for Requirements Prioritization Using Evolutionary Computing. Program Comput Soft 47, 746–756 (2021). https://doi.org/10.1134/S0361768821080235
[10]. K. Olmos-Sanchez, J. Rodas-Osollo, “Helping organizations manage the innovation process to join the Cognitive era”, Proc. - 2020 8th Ed. Int. Conf. Softw. Eng. Res. Innov. CONISOFT 2020, pp. 1–10, 2020, doi: 10.1109/CONISOFT50191.2020.00012.
[11]. I. Nonaka and H. Takeuchi, “The knowledge-creating company: how japanese companies create the dynamics of innovation”, in Hardvard Business Review, 1995, pp. 96–104.
[12]. K. Olmos-Sánchez, J. R. Osollo, L. F. Martínez, and V. M. Rocha, “Requirements engineering based on knowledge: A comparative case study of the KMoS-RE strategy and the DMS process”, Rev. Fac. Ing., vol. 2015, no. 77, pp. 88–94, 2015, doi: 10.17533/udea.redin.n77a11.
[13]. E. Morin, “From the Concept of System to the Paradigm of Complexity” J. Soc. Evol. Syst., vol. 15, no. 4, pp. 371–385, 1992.
[14]. P. Checkland, “Four Conditions for Serious Systems Thinking and Action”, Syst. Res. Behav. Sci., vol. 29, no. 5, pp. 465–469, 2012, doi: 10.1002/sres.2158.
[15]. P. Checkland and J. Poulter, “Soft Systems Methodology”, Syst. Approaches to Mak. Chang. A Pract. Guid., pp. 201–253, 2020, doi: 10.1007/978-1-4471-7472-1_5.
[16]. A. M. Jiménez-Galina, M. K. Olmos-Sánchez, A. A. Maldonado-Macías, J. Garnica-González, and J. G. Licona-Olmos, “Problem tree as a methodology to identify causes and effects of complex systems”, 2023 12th Int. Conf. Softw. Process. Improv. (CIMPS), 12 191–198. https://doi.org/10.1109/CIMPS61323.2023.10528840.

Информация об авторах / Information about authors

Алисия Маргарита ХИМЕНЕС-ГАЛИНА имеет степень магистра прикладных вычислений, с отличием окончила Автономный университет Сьюдад-Хуарес и в настоящее время готовится к защите диссертации для получения степени PhD в области передовых инженерных наук. Координатор по инновациям в Центре инноваций и интеграции передовых технологий (CIITA) Национального политехнического института Мексики (IPN). Область научных интересов: архитектура сложных систем, инновации в системах мониторинга окружающей среды, разработка заказного программного обеспечения и управление базами данных.

Alicia Margarita JIMÉNEZ GALINA, M.Sc. – Innovation Coordinator at the Center for Innovation and Integration of Advanced Technologies (CIITA) of the National Polytechnic Institute (IPN). She holds a Master's degree in Applied Computing, graduated with honors from the Autonomous University of Ciudad Juárez, and is currently pursuing a Ph.D. in Advanced Engineering Sciences. Research interests: complex systems architecture, innovation in environmental monitoring systems, custom software development, and database management.

Карла ОЛМОС-САНЧЕС закончила аспирантуру и получила степень PhD по инженерным наукам в Автономном университете Сьюдад-Хуарес (UACJ) по инженерии и управлению знаниями. Получила степень магистра в Национальном центре исследований и технологических разработок (CENIDET) по обработке естественного языка и входит в состав научного совета по прикладному искусственному интеллекту. При подготовке своей диссертации работала над методами управления знаниями о стратегии разработки требований (KMoS-RE), которая в первую очередь фокусируется на выявлении требований к знаниям в областях с неформальными структурами для концептуализации когнитивных решений. Эта стратегия была опубликована в различных книгах и международных журнальных статьях, и с 2015 года используется в прикладных вычислительных проектах при подготовке магистерских диссертаций. Имеет более чем 25-летний опыт преподавания предметов по программе компьютерной системной инженерии в университете UACJ. Руководила научной работой соискателей нескольких дипломных работ, магистерских и

докторских диссертаций. Координировала научную программу подготовки магистров прикладных вычислений и руководила концептуализацией, анализом и дизайном научной программы подготовки магистров прикладной кибербезопасности в UACJ.

Karla OLMOS-SÁNCHEZ completed her PhD studies in Engineering Sciences at the Autonomous University of Ciudad Juárez (UACJ) in Engineering and Knowledge Management. Additionally, she pursued her Master's degree at CENIDET (National Center for Research and Technological Development) in natural language processing and is currently part of the Applied Artificial Intelligence academic body. During her PhD, she developed the KMoS-RE strategy (Knowledge Management on a Strategy for Requirements Engineering), which primarily focuses on eliciting knowledge requirements in domains with informal structures to conceptualize cognitive solutions. This strategy has been published in various books, and international journal articles, and has been used in applied computing master's projects for thesis solution conceptualization since 2015. In her academic career, Dr. Olmos has over 25 years of teaching experience in computer science-related subjects in the Computer Systems Engineering program at UACJ. She has also supervised several undergraduate, master's, and doctoral theses. Additionally, she has coordinated the Applied Computing Master's program and led the conceptualization, analysis, and design of the Applied Cybersecurity Master's program at UACJ.

Аида Арасели МАЛЬДОНАДО-МАЧИАС получила степень PhD в области промышленной инженерии в Технологическом институте Сьюдад-Хуарес. В настоящее время она является профессором и научным сотрудником Автономного университета Сьюдад-Хуарес (UACJ). Сертифицированный профессиональный эргономист Коллегии эргономистов Мексики и действующим Президентом общества эргономистов. Член Национальной системы исследователей уровня II. Обладатель Государственной премии Мексики в области науки и техники 2018 года, награждена медалью штата Чиуауа 2020 года. Публиковала свои научные работы в нескольких индексируемых и известных журналах на английском и испанском языках, автор или соавтор более чем 100 статей и 10 книг на английском и испанском языках, в том числе во всемирно признанных издательствах. Рецензент и редактор известных журналов.

Aide Aracely MALDONADO-MACÍAS holds a Ph.D. in Industrial Engineering from the Instituto Tecnológico de Ciudad Juárez. She is currently a full-time Professor and Researcher at the Universidad Autónoma de Ciudad Juárez (UACJ). She is a Certified Professional Ergonomist by the Colegio de Ergonomistas de México and the current president of the Society of Ergonomists. She is also a member of the National System of Researchers as an SNI II. Dr. Maldonado received the 2018 State Science and Technology Award and the 2020 Distinguished Chihuahua Medal. She has published in multiple indexed and high-impact journals in both English and Spanish, with over 100 articles as an author or co-author and more than 10 books in English and Spanish. She is a reviewer and editor for prestigious journals and has also contributed to book chapters and authored books for internationally recognized publishers.

Жасмин Георгина ЛИКОНА-ОЛМОС, получила магистерскую степень по промышленной инженерии в Автономном университете мексиканского штата Идальго и степень PhD по системной инженерии в Национальном Политехническом институте. Она неоднократно публиковала результаты своих прикладных исследований по систематизации и сложным системам. Более 20 лет она ведет преподавательскую деятельность, осуществляя руководство студенческими работами.

Jazmín Georgina LICONA-OLMOS completed her PhD in Systems Engineering at the National Polytechnic Institute and her Master's studies in Industrial Engineering at the Autonomous University of the State of Hidalgo. She has several applied research publications in Systemics and

Complex Systems. In addition, she has an academic career of more than 20 years directing undergraduate and graduate theses.

Хулия Патрисия САНЧЕС-СОЛИС имеет степень PhD по программированию и с 2017 года является профессором Автономного университета Сьюдад-Хуарес. Ее научные интересы включают многоцелевую оптимизацию, машинное обучение и био-метаэвристики.

Julia Patricia SÁNCHEZ SOLIS holds a PhD in Computer Science and has been a full-time professor at the Autonomous University of Ciudad Juárez since 2017. Her research interests include multi-objective optimization, machine learning, and bio-inspired metaheuristics.

DOI: 10.15514/ISPRAS-2024-36(6)-9

Thematic Synthesis of Behavior-Driven Development: An Analytical Approach

V. M. Arredondo-Reyes, ORCID: 0009-0002-0218-8177 <vctmanuel.arrys@gmail.com>
S. Domínguez-Isidro, ORCID: 0000-0002-9546-8233 <sauldominguez@uv.mx>
Á. J. Sánchez-García, ORCID: 0000-0002-2917-2960 <angesanchez@uv.mx>
J. O. Ocharán-Hernández, ORCID: 0000-0002-2598-1445 <jocharan@uv.mx>
Faculty of Statistics and Informatics, Veracruz University, Xalapa, México.

Abstract. Behavior-driven development (BDD) focuses on specifying system behavior through examples, fostering collaboration, and aligning development with business needs. This research provides a thematic synthesis of BDD, highlighting its challenges, benefits, and implications in software development. By analyzing 23 studies across four academic databases, the study identifies trends and themes in BDD adoption and implementation. The findings emphasize BDD's role in bridging the gap between technical and non-technical stakeholders, aligning software development with business goals. Despite initial adoption challenges, the study reveals significant long-term benefits in software quality and stakeholder satisfaction. Future research should focus on developing efficient training and tools to support BDD adoption in diverse environments.

Keywords: behavior-driven development; software development; BDD adoption; BDD implementation; BDD challenges; BDD benefits; thematic synthesis.

For citation: Arredondo-Reyes V. M., Domínguez-Isidro S., Sánchez-García Á. J., Ocharán-Hernández J. O. Thematic synthesis of behavior-driven development: an analytical approach. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 161-178. DOI: 10.15514/ISPRAS-2024-3(6)-9.

Тематический синтез разработки, ориентированной на поведение: аналитический подход

В. М. Аррендондо-Рейес, ORCID: 0009-0002-0218-8177 <vctmanuel.arrys@gmail.com>
С. Домингес-Исидро, ORCID: 0000-0002-9546-8233 <sauldominguez@uv.mx>
А. Х. Санчес-Гарсия, ORCID: 0000-0002-2917-2960 <angesanchez@uv.mx>
Х. О. Очаран-Эрнандес, ORCID: 0000-0002-2598-1445 <jocharan@uv.mx>

Университет Веракруса, факультет статистики и информатики, Халапа, Мексика.

Аннотация. Разработка на основе поведения (BDD) фокусируется на определении поведения системы с помощью примеров, поощрений совместной работы и согласований разработки с потребностями бизнеса. В предлагаемой статье авторы описывают результаты изучения тематического синтеза BDD, подчеркивая его проблемы, преимущества и последствия для разработки программного обеспечения. Анализируя 23 исследования, ход которых отражен в четырех академических базах данных, исследование выявляет тенденции и направления в следовании принципам и реализации BDD. Авторами подчеркивается роль BDD в преодолении разрыва между техническими и нетехническими заинтересованными сторонами, согласовании разработки программного обеспечения с бизнес-целями. Несмотря на первоначальные проблемы с внедрением BDD, проведенное исследование показывает его значительное долгосрочное и благотворное влияние на качество программного обеспечения, а также на достижение удовлетворенности заинтересованных сторон. Будущие исследования должны быть сосредоточены на разработке эффективного обучения и инструментов для поддержки внедрения BDD в различных средах.

Ключевые слова: разработка, основанная на поведении (BDD); разработка программного обеспечения; внедрение BDD; реализация BDD; проблемы BDD; преимущества BDD; тематический синтез.

Для цитирования: Аррендондо-Рейес В. М., Домингес-Исидро С., Санчес-Гарсия А. Х., Очаран-Эрнандес Х. О. Тематический синтез разработки, ориентированной на поведение: аналитический подход. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 161–178 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–9.

1. Introduction

This paper builds on the research presented at the 2023 11th International Conference on Software Engineering Research and Innovation (CONISOFT) [1] by conducting a Systematic Literature Review (SLR) to analyze the challenges and benefits of Behavior-Driven Development (BDD). BDD, an evolution of Test-Driven Development (TDD), is a significant advancement in software development, focusing on defining and developing software based on system behavior rather than solely verifying functionality through tests [2]. This characteristic enables software development teams to focus on identifying, understanding, and subsequently building valuable features that interest businesses, ensuring they are implemented effectively [3].

BDD, with its proactive and collaborative approach, has significantly impacted the industry in recent years as teams strive to deliver high- quality software that meets stakeholder requirements [4]. By collaboratively defining expected system behavior at the outset, BDD allows early identification of potential issues, preventing costly misunderstandings and rework [5]. Furthermore, BDD promotes the development of automated tests that verify software behavior, aiding in the detection of regressions and ensuring that new features do not compromise existing functionality.

Since BDD continues to be adopted across the industry, understanding its benefits and challenges is crucial for successful implementation. In this context, we extend our previous analysis [1] in order to contribute to this understanding by offering a thematic synthesis of BDD's application, highlighting critical factors for its effective adoption, and providing recommendations to address common challenges. This thematic synthesis identifies trends and common themes across the studies analyzed, offering a documented overview of BDD's implementation.

The rest of the document is organized as follows: Section 2 reviews related work; Section 3 describes the materials and research methods; Section 4 summarizes the systematic review results. Section 5 discusses thematic synthesis. Section 6 addresses validity threats, and finally, Section 7 presents conclusions and future work.

2. Related work

The software industry has significantly evolved, integrating automatic data processing through ICTs into various social niches. As software demand grows, maintaining high- quality products is crucial, prompting the adoption of methods for better design, implementation, and maintenance of software systems [6]. This need has driven studies exploring the implications of these methods.

Three relevant studies were identified. Abushama et al. [7] systematically reviewed the impact of TDD and BDD on project success factors such as cost, time, and customer satisfaction, analyzing 31 studies. Their findings suggest that while BDD may incur higher costs and time, it tends to achieve greater customer satisfaction. Arnyndiasari et al. [8] reviewed Agile methodologies, including BDD, highlighting the benefits of integrating these practices to enhance development success. Farooq et al. [9] focused on BDD, emphasizing its role in clarifying requirements and bridging communication gaps between stakeholders and developers, leading to higher customer satisfaction. However, they noted potential challenges in BDD implementation.

Our study differs by analyzing software project environments where BDD has been implemented and identifying critical insights into its adoption, challenges, and utility.

Table 1. Comparison of related works.

Characteristic	Related works		
	[7]	[8]	[9]
Year	2020	2022	2023
Approach	Analysis of the impact of TDD and BDD on time, cost, and customer satisfaction.	Review Agile methodologies (TDD, BDD, DDD, MDD) and their effectiveness.	Evaluation of BDD and its impact on software development and product quality.
Research Questions	Impact of TDD and BDD on project success factors.	Characteristics and effects of TDD, BDD, DDD, and MDD.	Techniques to reduce ambiguities and communication gaps in BDD.
Findings	BDD achieves higher customer satisfaction compared to TDD; more research is necessary.	Integrating Agile methodologies can improve software development success.	BDD is effective in clarifying requirements and improving communication between stakeholders and developers.
Coverage	Systematic literature review (1999-2020, 31 studies).	Systematic literature review (2000-2021, 16 studies)	Systematic literature review (2010-2022, 31 studies) Development of framework and taxonomy for BDD.

3. Materials and method

This research followed the Systematic Literature Review (SLR) guidelines by Kitchenham et al. [10]. The method consists of three main stages: 1) planning, 2) Identifying the state-of-the-art in BDD, and 3) Interpreting the results. For the last stage, the method provided by Popay et al. [11] for narrative synthesis for the SLR was applied. The development of thematic synthesis was based on the process of Cruzes and Dybå [12]. Finally, to conduct the thematic synthesis, we based on the guide by Uştuk [13] on a thematic synthesis with MAXQDA.

3.1 Research Planning

The state-of-the-art analysis is guided by five research questions (RQs) aimed at i) Identifying and describing the characteristics of projects where BDD has been successfully implemented. ii) Determining the scenarios where BDD benefits software development. iii) Identifying challenges related to BDD implementation and how they can be addressed. iv) Specifying the knowledge needed to enhance BDD's adaptability, facilitating its adoption across different environments and teams. v) Documenting the advantages BDD brings to software projects, such as improved product quality and stakeholder satisfaction.

1. **RQ1.-** What are the characteristics of the projects in which BDD has been used to develop software?
2. **RQ2.-** What are the specific scenarios in which BDD benefits software development?
3. **RQ3.-** What are the challenges in the use of BDD?
4. **RQ4.-** What information should be known to increase the degree of adaptability of BDD?
5. **RQ5.-** What are the reported benefits of using BDD?

Concerning the search strategy, a search string was generated through an elicitation process presented in [10], resulting in the following:

("behavior driven development" OR "behavior-driven development" OR "behavioural-driven development") AND (tendencies OR benefits OR advantage OR trends)

Information sources for the automated search include four key databases: IEEE Xplore, ACM Digital Library, SpringerLink, and Science Direct, which store relevant proceedings and journal papers in software engineering and related fields.

The study selection process is divided into five phases, applying inclusion (IC) and exclusion criteria (EC). Phase 1 includes studies published between 2015 and 2023 (IC1) and written in English (IC2). Studies before 2015, book chapters, monographs, theses (EC1), and secondary studies (EC2) are excluded. Phase 2 includes papers with relevant search terms in their title, abstract, or keywords (IC3) and excludes demos or inaccessible works (EC3). In Phase 3, studies unrelated to software development (EC4) are excluded, while those with abstracts related to research questions (IC4), documented results (IC5), and published in selected sources (IC6) are included. Phase 4 excludes duplicates (EC5). Finally, Phase 5 includes works that directly answer a research question (IC7). This selection process is applied to both automated search and snowballing.

In order to measure the relevance and impact of the selected study on this research, we applied a seven-question checklist based on the criteria shown by Dybå and Dingsøyr [14].

1. **Q1:** Is the document based on research, or is it a "lessons learned" report based on an expert opinion?
2. **Q2:** Is there an explicit statement of research objectives?
3. **Q3:** Is there a sufficient description of the context in which the proposed methodology was tested?
4. **Q4:** Does the research design address the objectives adequately?
5. **Q5:** Was information obtained that addressed characteristics of the project in which the methodology was used?
6. **Q6:** Does the study provide value in research or practice?
7. **Q7:** Was the proposal for using the methodology evaluated?

The scores assigned to the papers reflect the quality of their contributions to this research. Papers are ranked based on their final score, ranging from 7 to 5. A score of 7 or higher is considered high-ranked, a score between 4 and less than six is considered average, and any score below four is deemed low-ranked.

4. Results

This section provides a summary of the results from the systematic literature review (SLR) presented in [1], as well as an overview of the thematic synthesis derived from the analysis

4.1 Selection of Primary Studies

We got 371 publications from the selected databases using the proposed search string. Twenty-five articles were selected after applying the inclusion and exclusion criteria to the databases, see Fig. 1.

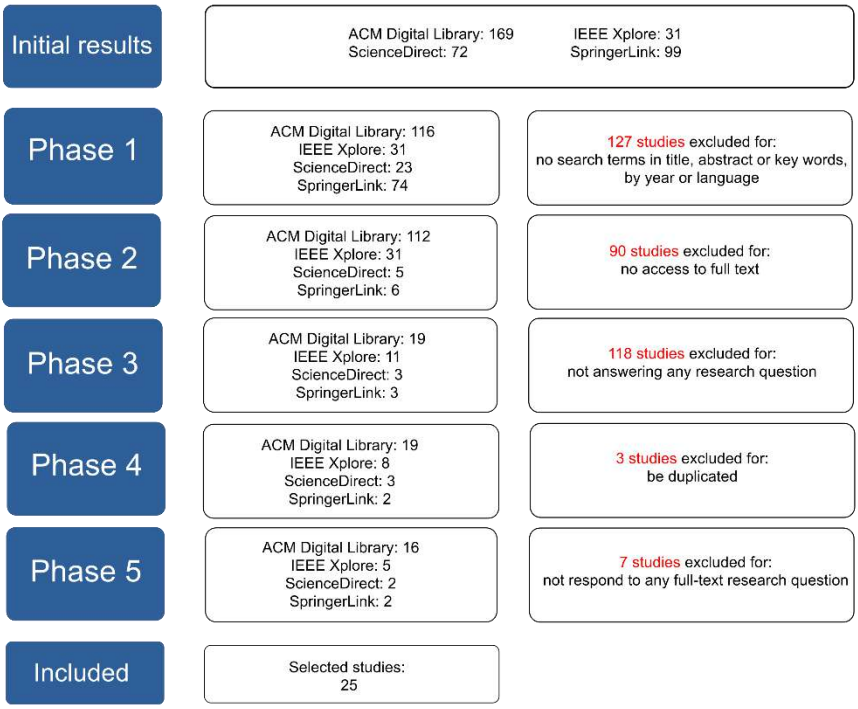


Fig. 1. Selection phases for automatic search.

Wohlin et al. [15] described the snowballing technique, which expanded the study pool by adding 550 backward and 158 forward studies. After applying the selection criteria, only three additional studies were included, resulting in 28 primary studies. Table 2 lists these selected studies and their quality scores based on questions Q1–Q6.

4.2 Studies Distribution

Behavior-driven development (BDD) research has increased significantly since 2018, with 54% of the work concentrated between 2018 and 2020 and 35% in the last three years. In the last two years, studies increased by 60% (see Fig. 2).

The International Conference on Software Engineering (ICSE) is the primary venue for publishing articles on BDD, accounting for 17% of the selected studies. ICSE is renowned for its comprehensive coverage of software engineering topics and attracts leading experts, fostering collaboration and knowledge exchange. The International Conference Proceedings Series (ICPS) follows, accounting for 13% of the selected studies. ICPS is recognized for its interdisciplinary approach, making it an attractive venue for BDD research due to its encouraging cross-pollination

of ideas and perspectives. The Brazilian Symposium on Software Engineering also emerges as a significant venue, hosting 8% of the selected studies. Although not as widely recognized as ICSE or ICPS, it provides a valuable platform for regional researchers to contribute to the discourse on BDD.

Table 2. Studies selected ordered by quality.

Reference	Year	Data base	Quality score
[16]	2021	Science Direct	7
[17]	2018	ACM Digital Library	7
[18]	2018	IEEE Xplore	7
[19]	2023	SpringerLink	7
[20]	2018	ACM Digital Library	7
[21]	2019	ACM Digital Library	6
[22]	2019	ACM Digital Library	6
[23]	2020	ACM Digital Library	6
[24]	2020	ACM Digital Library	6
[25]	2016	SpringerLink	6
[26]	2023	IEEE Xplore	6
[27]	2023	ACM Digital Library	6
[28]	2015	IEEE Xplore	5
[29]	2022	Science Direct	5
[30]	2018	ACM Digital Library	5
[31]	2020	ACM Digital Library	5
[32]	2020	IEEE Xplore	5
[33]	2021	ACM Digital Library	5
[34]	2022	IEEE Xplore	5
[35]	2020	SpringerLink	5
[36]	2020	ACM Digital Library	5
[37]	2019	ACM Digital Library	5
[38]	2021	IEEE Xplore	5
[39]	2023	ACM Digital Library	5
[40]	2023	ACM Digital Library	5
[41]	2018	IEEE Xplore	4
[42]	2017	ACM Digital Library	4
[5]	2018	ACM Digital Library	3

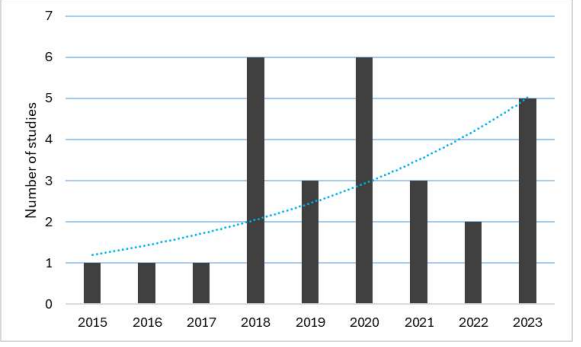


Fig. 2. Selected studies by year. The dashed line indicates an upward trend.

4.3 Thematic Synthesis

The thematic synthesis followed the 21-step approach by Cruzes and Dybå [12]. The process began with information understanding, where the 23 selected studies were thoroughly reviewed to identify relevant text segments. These segments were labeled and coded, resulting in 51 initial codes. After peer review, the list was refined to 27 codes, which were then translated into five cohesive themes.

1. **Development Aspects**: Explores project characteristics influencing BDD implementation, addressing RQ1 and RQ2.
2. **Benefits**: Focuses on BDD's positive impacts on the development cycle, addressing RQ5.
3. **Best Practices**: Delves into effective BDD implementation practices aligned with RQ4.
4. **Difficulties**: Examines challenges in BDD usage corresponding to RQ3.
5. **Usage Recommendations**: Offers expert guidance for BDD implementation, also addressing RQ4.

The thematic map (Fig. 3) illustrates the hierarchical organization of these themes and subthemes, effectively answering the research questions. It categorizes findings into key domains such as Development Aspects, Benefits, Best Practices, Recommendations, and Challenges.

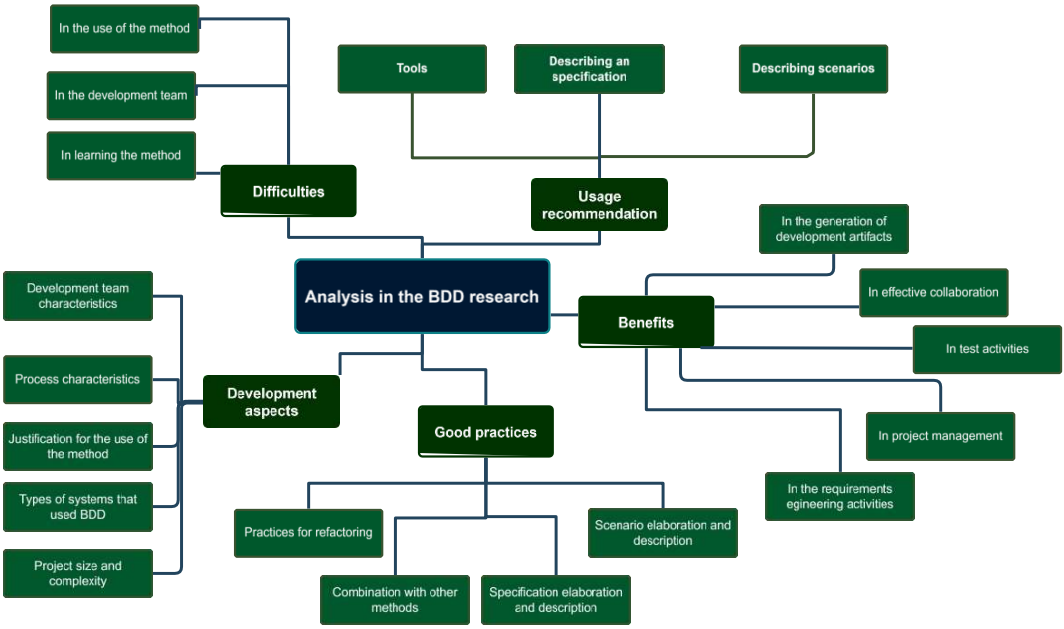


Fig. 3. Thematic map proposed.

5. Discussion and analysis

Below are the five main BDD themes identified in our research.

5.1 Development Aspects

The use of BDD in software projects encompasses various characteristics, from team details to process features, the justification for its adoption, project size, and the types of systems developed. These aspects are explored in detail below.

5.1.1 Development Team Characteristics

Successful BDD implementation depends on team characteristics. In large-scale and geographically distributed teams, BDD enhances communication and coordination [16, 22]. In educational settings, writing acceptance tests before development integrates quality early [17, 21]. The team's prior experience also affects BDD's efficiency and productivity [21, 28].

5.1.2 Process Characteristics

BDD's application varies by context. In distributed systems, it emphasizes reusing scenarios and test steps [16, 29, 30]. In agile projects, BDD improves quality, organization, and collaboration [23, 31]. During requirements validation, BDD enhances organization and collaboration throughout development [23].

5.1.3 Justification for Adopting BDD

BDD is adopted in large-scale projects to address challenges like team coordination and requirements management. It facilitates requirement documentation and coordination in telecommunications and enterprise systems [16, 24, 29]. In geographically distributed projects, BDD mitigates communication barriers and fosters a common language [5, 22]. For critical systems, such as in the automotive industry, BDD improves requirement specification and system validation [24, 34].

5.1.4 Project Size and Complexity.

BDD’s effectiveness varies with project size and complexity. While initially suited for medium-sized projects, it also benefits larger projects [32]. However, extensive BDD test suites can increase maintenance and comprehension complexity [35].

5.1.5 Types of Systems Using BDD

BDD is applied across various sectors and technologies. It is used in complex projects like telecommunications and microservices architectures [16, 29]. In critical systems, such as automotive, BDD ensures system integrity and reliability [24, 28, 34]. It also adapts to emerging technologies and diverse development paradigms [21, 32, 42].

5.2 Benefits

This subsection covers how BDD enhances various aspects of the development cycle, including requirements management and the creation of valuable artifacts. It also improves quality, stakeholder communication, and critical phases like requirements and testing.

5.2.1 Effective Collaboration

BDD enhances collaboration and communication between quality engineers and business analysts. Scenario refactoring improves interaction, and clear visualization of test scenarios aids comprehension even for non-technical stakeholders [29, 32]. Precise language in behavior scenarios reduces misunderstandings, fosters better project management, and improves communication in geographically distributed teams [5, 36, 41]. BDD also integrates clients more closely through scenario-based documentation, enhancing collaboration and product quality [18, 24, 28].

5.2.2 Artifact Generation

BDD improves artifact generation by expressing requirements as executable test cases, reducing rework and saving time [16, 21]. It provides automated acceptance tests, enhances requirement elicitation, and creates “living documentation” that evolves with the system [18, 23, 24, 25, 28]. The approach also facilitates test case reusability, benefiting development and verification [16, 18].

5.2.3 Testing Activities

BDD simplifies and structures testing activities, improving efficiency and automation. It organizes tests by features and scenarios, aids maintenance, and enhances test completeness and stability [17, 18, 22]. BDD improves predictability and confidence in code stability by defining behaviors before implementation and automating tests [18, 31, 41]. Overall, BDD enhances test efficiency, completeness, and process ease [5, 17, 19, 20, 33].

5.2.4 Requirement Engineering Activities

BDD enhances the quality and understanding of requirements by expressing them as executable test cases. The Given-When- Then format clarifies business perspectives and improves requirement clarity [16, 17]. BDD facilitates discussions, improves traceability between requirements and code, and creates "living documentation" [5, 23, 33]. It also addresses security requirements and reduces ambiguities [16, 17, 28, 42].

5.2.5 Project Management

BDD supports project management by improving scenario grouping reducing development time and costs. Early scenario development enhances efficiency, and "living documentation" provides continuous updates [16, 18]. BDD improves code quality and productivity, benefiting exploratory testing and product quality [18, 41].

5.3 Best Practices

This section provides practical guidelines for applying BDD, covering maintainable specifications, new specification creation, and scenario refactoring techniques.

5.3.1 Refactoring Practices

Refactoring is essential for improving software quality and maintainability. Key practices include [29]:

Identification of Areas for Refactoring

- **Preprocessing:** Store each BDD specification in a separate file with the name on the first line and steps on subsequent lines. Remove BDD keywords for easier comparison.
- **Measurement:** Using automated scripts, calculate normalized compression similarity (NCS) and Similarity Ratio (SR) for all specification pairs.
- **Ranking:** Analyze and rank the NCS and SR values to determine a similarity between specifications.

Careful Application of Refactoring

- **Merging:** Combine specifications with common lines and minimal differences.
- **Restructuring:** Create new statements from common statements with different specifications.
- **Deleting:** Remove duplicate or functionally identical specifications.
- **Renaming:** Rename specifications with similar names but different functionalities to avoid ambiguity.

Validation to Preserve Behavior

- Ensure that refactoring does not alter the software's behavior.

5.3.2 Combination with Other Methods

Combining BDD with other methods can enhance communication, address complexity, and reduce gaps in distributed teams. Benefits include improved security verification, communication, and reduced inconsistencies in early development phases [17, 28, 32, 42]. Various studies suggest integrating BDD with other techniques to improve development quality.

5.3.3 Elaboration and Description of Specifications

Structural Practices

- Develop system-level feature files and hooks for effective integration testing [16, 35].

Practices for Developing New BDD Specifications

- **Specify New Behaviors:** Product managers should write new behaviors based on customer requests.
- **Develop System-Level Feature Files:** Create detailed, executable feature files that outline approved behaviors [16].

Practices Related to SBVR and Event-B

- **Determine Business Objectives:** Collaborate with clients and analysts to establish project objectives [42].
- **Define Software Functionalities:** Refine goals into a list of features with a specific format [42].
- **Define Acceptance Criteria:** Create scenarios representing acceptance criteria using the given-when-then format [29].

Improvement Areas

- Regular feedback is crucial for early correction and alignment with team objectives [16].

5.3.4. Elaboration and Description of Scenarios

We identified four principal pieces of information related to writing scenarios BDD: formal redaction, simplifying scenarios, evading ambiguity, and establishing a limit for the step in a scenario to adequate comprehension.

Key Information for Writing BDD Scenarios

- **Abstraction Level:** Maintain an appropriate level of detail to balance understanding and code complexity [35].
- **Reuse of Step Phrases:** Use existing steps to enhance readability and maintainability [32, 39].
- **Balance Generic and Specific Steps:** Combine generic steps with parameters and specific steps to improve readability and execution. Reusing steps with parameters and generic names like "When the user clicks on the '<element name>' element on the '<page name>' page" is helpful. [32, 39].
- **Limit Actions in Scenarios:** Each scenario can have only a single "When" action. Split scenarios with multiple actions or move extra actions to the "Given" section [32, 39].
- **Indent "And" Steps:** Use "And" steps for improved readability [32].
- **Seek Reusable Behaviors:** Avoid redundant development and testing of similar behaviors [16].
- **Address Duplication:** Automate duplicate searches, refactor code frequently and adhere to the Single Responsibility Principle (SRP) [18].

Scenario Structuring Recommendations

- **Naming Pattern:** Use the "When..., then..." pattern for concise scenario descriptions [26].
- **Step Count:** Limit steps per scenario to 12 for better readability [26].
- **Step Order:** Follow the "Given", "When", and "Then" order. Multiple "When-Then" combinations may indicate the need for separate scenarios [26].
- **Perspective:** Write scenarios from a third-person perspective to avoid ambiguity, e.g., "When the user clicks on the button" instead of "When I click on the button" [26].
- **Domain Vocabulary:** Use precise terms and avoid duplicates for clarity. Minimize technical jargon to ensure all stakeholders understand [35].

5.4 Difficulties

Challenges in adopting the BDD method can be categorized into difficulties related to learning the method, development team issues, and the practical use of BDD. These challenges are further detailed below:

5.4.1 Learning the Method

The high learning curve associated with BDD presents significant obstacles. Resistance to BDD may arise from a lack of testing culture, as BDD requires a shift in perspective. The process's lack of visual appeal can also hinder adoption [24]. The steep learning curve is often exacerbated by limited experience with BDD, leading to initial difficulties and resistance, particularly in teams unfamiliar with the method [31]. These issues highlight the need for a supportive culture and thorough training to ease the transition to BDD [16, 29, 41].

5.4.2. Development Team Challenges

The lack of experience and commitment within the development team significantly impacts the successful adoption of BDD. Inexperienced team members may struggle with proper scenario specification, leading to issues such as scenario duplication and incomplete scenarios [18]. Communication and collaboration are also hindered by a lack of commitment, which is crucial for the success of agile methodologies [41]. The absence of method knowledge among team members further exacerbates these challenges, making it difficult to effectively implement and maintain BDD practices [23, 31].

5.4.3. Using the Method

Implementing BDD presents several practical challenges:

- **Scenario Management:** Managing scenarios in large-scale environments is complex due to the dynamic nature of requirements and the need for frequent iterations with domain experts. Maintaining an accurate record of behavior changes across multiple stakeholders adds to the complexity [16].
- **Adapting to New Environments or Requirements:** Modifying BDD specifications to reflect new business policies or environments can be challenging, particularly in large-scale projects. Updating BDD frameworks or adapting them to new requirements may result in duplicated efforts and slow down development [18].
- **Adopting BDD Tools and Technologies:** Introducing new BDD tools in large projects requires significant time and effort. Training is essential for achieving productivity in a BDD environment. Additionally, challenges arise when updating BDD frameworks across multiple microservices, requiring careful evaluation of tool suitability [23, 28, 41].

- **Comprehending BDD Specifications:** Understanding BDD specifications can be difficult, particularly when duplication in specifications hinders comprehension and unnecessarily prolongs test suite execution [18].
- **Specifying Scenarios:** Scenario specification in BDD is complex, especially in large-scale projects where requirements evolve over time. The lack of initial clarity and the need for frequent iterations complicate scenario specification, making it a challenging task [16, 23, 30, 31].
- **Specification Size:** Large projects pose additional challenges due to the exponential growth of possible scenarios. The complexity of managing and maintaining these scenarios can be overwhelming [36].
- **Maintainability of BDD Specifications:** The maintenance of BDD specifications is particularly challenging in large-scale projects. The high cost and complexity of maintaining these specifications can deter teams from adopting automated acceptance testing [29]. Effective maintenance strategies, such as refactoring, are necessary to manage the growing complexity and ensure the long-term success of BDD [16, 18, 35].

5.5 Usage Recommendations

Implementing BDD goes beyond adopting tools and practices; it requires understanding the guidelines and best practices for specifying requirements and crafting scenarios. This section provides key recommendations from experienced practitioners, divided into three areas: specification description, scenario elaboration, and tool usage.

5.5.1 For the Description of a Specification

The specification in BDD serves as a document that describes the desired system behaviors from a high-level perspective. It communicates how the software should meet requirements in natural language.

- **Limit actions per scenario**: Restricting the number of actions in each scenario maintains clarity and conciseness. This practice ensures that both technical and non-technical stakeholders can quickly grasp the system's functionality without unnecessary complexity [27, 32, 37].
- **Preserve domain vocabulary**: Using consistent domain-specific terms promotes shared understanding among teams, enhancing collaboration and ensuring alignment on the system's goals and requirements [35, 37].
- **Conserve a few steps**: Focus on essential steps to maintain clarity and conciseness in each scenario. This approach ensures scenarios remain understandable, especially for those not directly involved in development [32, 35].
- **Eliminate technical vocabulary**: Avoiding technical terms makes specifications accessible to all stakeholders, facilitating effective communication at the initial stage [35].

5.5.2. For the Description of Scenarios

Scenarios are concrete instances that exemplify how the system should behave in specific contexts.

- **Each scenario tests one thing**: Focus each scenario on testing a specific functionality or behavior, making it easier to identify issues during execution [35, 40].
- **Make descriptive titles**: Clear and descriptive titles help quickly identify the purpose of each test case [35].
- **Oriented towards customer benefit**: Write scenarios from the perspective of the benefit they offer to the end user, ensuring alignment with customer expectations [35].

- **Make an explicit and verifiable description**: Scenarios should be concise, clear, and easily verifiable, facilitating execution and ensuring understandable results [37].
- **Maintain singularity in scenarios**: Each scenario should clearly contribute to the overall quality assurance of the software, ensuring each test case adds value [37].
- **Avoid ambiguities**: Clarity is key; avoiding ambiguities ensures reliable test results with no room for misinterpretation [37].

5.5.3. Tools

Selecting and using tools in BDD is critical. Various studies highlight popular open-source tools such as Cucumber, Concordion, JBehave, FitNesse, and SpecFlow, recognized for their role in facilitating BDD processes [16, 29, 32, 41].

- **Obsolete Tools**: Some tools, including StoryQ, JDave, NBehave, Easyb, and BDDfy, are no longer actively maintained, underscoring the importance of choosing up- to-date tools with active community support [38].
- **Documentation Evaluation**: Clear and comprehensive documentation is vital for efficient adoption and learning, allowing teams to maximize tool capabilities [29, 38].
- **Consideration of Reference Projects**: Reviewing reference projects that use the selected tools can provide practical insights and improve BDD implementation [18].

Selecting tools for development is important; when using BDD, technical functionality, currency, documentation, and an active user community must be considered. Evaluating IDE plugins provides valuable information on how tools facilitate collaboration and behavior specification within the development environment.

6. Validity threats

We acknowledge potential threats to the validity of our results but have taken measures to mitigate them. One potential threat involves the study search and selection process, which relies on the researcher's judgment and includes non- English languages [43-44]. To address this, we adhered to guidelines by Kitchenham et al. [10]. Peer reviews were conducted by at least three authors, following the coding and theming process described by Cruzes and Dybå [12]. Additionally, we utilized the MAXQDA tool for thematic synthesis [13]. To ensure the relevance of selected studies, we employed the snowballing method [15], conducting one forward and one backward iteration. While limitations, such as excluding studies due to restricted access, are recognized, our findings offer a comprehensive understanding of BDD's applications, benefits, and challenges. Our aim is not to provide prescriptive guidance or solutions but to enlighten and inform.

7. Conclusion

This research delved into essential aspects of BDD, focusing on its principles, differences from other methodologies, and practical applications through a systematic literature review. Key conclusions include insights into BDD's applications, benefits, and challenges, as well as the identification of recommended practices and common difficulties. While the study provided valuable perspectives, it is important to acknowledge limitations, such as the reliance on existing studies and gray literature, highlighting the need for further investigation.

The systematic review revealed that BDD enhances communication, collaboration, and adaptability while minimizing requirements misunderstandings. It also identified trends in BDD’s application, including its benefits for collaboration, testing, requirements engineering, and project management.

Our research methodology involved a thorough systematic review, with a carefully tailored search strategy and quality assessments to ensure reliability. This comprehensive approach offers a robust foundation for understanding BDD's implementation, challenges, advantages, and best practices.

In conclusion, this study has significantly contributed to the understanding of BDD, underscoring its contemporary relevance and growing interest in the software development community. It offers valuable insights for those considering the adoption of behavior-driven agile methodologies, promoting the creation of well-designed, precisely adapted software solutions.

References

[1]. V. M. Arredondo-Reyes, S. Domínguez-Isidro, A. J. Sánchez-García, and J. O. Ocharán-Hernández, "Benefits and Challenges of the Behavior-Driven Development: A Systematic literature review," in 11th International Conference in Software Engineering Research and Innovation (CONISOFT 2023), IEEE, 2023, pp. 1–6.

[2]. M. M. Moe and J. C. Sanchez, "International Journal of Trend in Scientific Research and Development (IJTSRD) Comparative Study of Test-Driven Development (TDD), Behavior-Driven Development (BDD) and Acceptance Test-Driven Development (ATDD) the Creative Commons Attribution License (CC BY 4.0)." [Online]. Available: http://creativecommons.org/licenses/by/4.0

[3]. J. Ferguson and D. North, BDD IN ACTION Behavior-Driven Development for the whole software lifecycle. [Online]. Available: www.it-ebooks.info

[4]. C. Solís and X. Wang, "A study of the characteristics of behaviour driven development," in Proceedings - 37th EUROMICRO Conference on Software Engineering and Advanced Applications, SEAA 2011, pp. 383–387. doi: 10.1109/SEAA.2011.76.

[5]. L. Pereira, H. Sharp, C. De Souza, G. Oliveira, S. Marczak, and R. Bastos, "Behavior-driven development benefits and challenges: Reports from an industrial study," in Proceedings of the 19th International Conference on Agile Software Development: Companion, Association for Computing Machinery, 2018, pp. 1–4. doi: 10.1145/3234152.3234167.

[6]. "What is software development? IBM." [Online]. Available: https://www.ibm.com/topics/software-development

[7]. H. M. Abushama, H. A. Alassam, and F. A. Elhaj, "The effect of Test-Driven Development and Behavior-Driven Development on Project Success Factors: A Systematic Literature Review Based Study," in Proceedings of: 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering, ICCCEEE 2020, Institute of Electrical and Electronics Engineers Inc., Feb. 2021. doi: 10.1109/ICCCEEE49695.2021.9429593.

[8]. D. Arnyndiasari, R. Ferdiana, and P. I. Santosa, "Software Practices for Agile Developers: A Systematic Literature Review," in 2022 1st International Conference on Information System and Information Technology, ICISIT 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 238–243. doi: 10.1109/ICISIT54091.2022.9872874.

[9]. M. S. Farooq, U. Omer, A. Ramzan, M. A. Rasheed, and Z. Atal, "Behavior Driven Development: A Systematic Literature Review," IEEE Access, vol. 11, pp. 88008–88024, 2023, doi: 10.1109/ACCESS.2023.3302356.

[10]. B. A. Kitchenham, D. Budgen, and P. Brereton, Evidence-Based Software Engineering and Systematic Reviews. Chapman & Hall/CRC, 2015.

[11]. J. Popay et al., "Guidance on the Conduct of Narrative Synthesis in Systematic Reviews," 2006.

[12]. D. S. Cruzes and T. Dybå, "Recommended Steps for Thematic Synthesis in Software Engineering," in 2011 International Symposium on Empirical Software Engineering and Measurement, IEEE, 2011, pp. 275-284. doi: 10.1109/ESEM.2011.36.

[13]. Ö. Uştuk, "Thematic Analysis with MAXQDA: Step- by-Step Guide," MAXQDA. [Online]. Available: https://www.maxqda.com/blogpost/thematic-analysis- with-maxqda-step-by-step-guide.

[14]. T. Dybå and T. Dingsøyr, "Empirical studies of agile software development: A systematic review," Inf Softw Technol, vol. 50, no. 9, pp. 833–859, 2008, doi: 10.1016/j.infsof.2008.01.006.

[15]. C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, ACM, 2014, pp. 1–10. doi: 10.1145/2601248.2601268.

[16]. M. Irshad, R. Britto, and K. Petersen, "Adapting Behavior Driven Development (BDD) for large-scale software systems," Journal of Systems and Software, vol. 177, no. 110944, pp. 1–20, 2021, doi: 10.1016/j.jss.2021.110944.

[17]. H. Bünder and H. Kuchen, "Towards Behavior-Driven Graphical User Interface Testing," in Proceedings of the ACM Symposium on Applied Computing, Association for Computing Machinery, 2019, pp. 1742–1751. doi: 10.1145/3297280.3297450.

[18]. Y. Wang and S. Wagner, “Combining STPA and BDD for safety analysis and verification in agile development: A controlled experiment,” in Agile Processes in Software Engineering and Extreme Programming. XP 2018. Lecture Notes in Business Information Processing, Springer, Cham, 2018, pp. 37–53. doi: 10.1007/978-3-319-91602-6_3.

[19]. A. Scandaroli, R. Leite, A. H. Kiosia, and S. A. Coelho, “Behavior-Driven Development as an Approach to Improve Software Quality and Communication across Remote Business Stakeholders, Developers and QA: Two Case Studies,” in 2019 ACM/IEEE 14th International Conference on Global Software Engineering (ICGSE), 2019, pp. 105-110. doi: 10.1109/ICGSE.2019.00030.

[20]. M. Rahman and J. Gao, “A reusable automated acceptance testing architecture for microservices in behavior-driven development,” in 9th IEEE International Symposium on Service-Oriented System Engineering, IEEE SOSE 2015, Institute of Electrical and Electronics Engineers Inc., 2015, pp. 321–325. doi: 10.1109/SOSE.2015.55.

[21]. M. Irshad, J. Börstler, and K. Petersen, “Supporting refactoring of BDD specifications—An empirical study,” Inf Softw Technol, vol. 141, no. 106717, pp. 1–13, 2022, doi: 10.1016/j.infsof.2021.106717.

[22]. A. Dimanidis, K. C. Chatzidimitriou, and A. L. Symeonidis, “A Natural Language Driven Approach for Automated Web API Development: Gherkin2OAS,” in Companion Proceedings of the The Web Conference 2018, Association for Computing Machinery, 2018, pp. 1869–1874. doi: 10.1145/3184558.3191654.

[23]. N. Nascimento, A. R. Santos, A. Sales, and R. Chanin, “Behavior-Driven Development: An Expert Panel to evaluate benefits and challenges,” in Proceedings of the XXXIV Brazilian Symposium on Software Engineering, Association for Computing Machinery, 2020, pp. 41–46. doi: 10.1145/3422392.3422460.

[24]. N. Nascimento, A. R. Santos, A. Sales, and R. Chanin, “Behavior-Driven Development: A case study on its impacts on agile development teams,” in Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops, Association for Computing Machinery, 2020, pp. 109–116. doi: 10.1145/3387940.3391480.

[25]. R. K. Lenka, S. Kumar, and S. Mamgain, “Behavior Driven Development: Tools and Challenges,” in Proceedings - IEEE 2018 International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2018, Institute of Electrical and Electronics Engineers Inc., pp. 1032–1037. doi: 10.1109/ICACCCN.2018.8748595.

[26]. O. Bezsmertnyi, N. Golian, V. Golian, and I. Afanasieva, “Behavior Driven Development Approach in the Modern Quality Control Process,” in 2020 IEEE International Conference on Problems of Infocommunications Science and Technology, PIC S and T 2020 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 217–220. doi: 10.1109/PICST51311.2020.9467891.

[27]. T. R. Silva and B. Fitzgerald, “Empirical findings on bdd story parsing to support consistency assurance between requirements and artifacts,” in Proceedings of the 25th International Conference on Evaluation and Assessment in Software Engineering, Association for Computing Machinery, 2021, pp. 266–271. doi: 10.1145/3463274.3463807.

[28]. C. Wiecher, S. Japs, L. Kaiser, J. Greenyer, R. Dumitrescu, and C. Wolff, “Scenarios in the loop: Integrated requirements analysis and automotive system validation,” in 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems, MODELS-C 2020 - Companion Proceedings, Association for Computing Machinery, Inc, 2020, pp. 199–208. doi: 10.1145/3417990.3421264.

[29]. C. Lauer and C. Sippl, “Benefits of Behavior Driven Development in Scenario-based Verification of Automated Driving,” in 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Institute of Electrical and Electronics Engineers (IEEE), 2022, pp. 105–110. doi: 10.1109/itsc55140.2022.9922498.

[30]. F. L. Siqueira, T. C. de Sousa, and P. S. Silva, “Using BDD and SBVR to Refine Business Goals into an Event-B Model: A Research Idea,” in 2017 IEEE/ACM 5th International FME Workshop on Formal Methods in Software Engineering (FormaliSE), 2017, pp. 31–36. doi: 10.1109/FormaliSE.2017.5.

[31]. T. R. Silva, J. L. Hak, and M. Winckler, “Testing Prototypes and Final User Interfaces Through an Ontological Perspective for Behavior-Driven Development,” in Bogdan, C., et al. Human-Centered and Error-Resilient Systems Development. HESSD HCSE 2016 2016. Lecture Notes in Computer Science, Springer, Cham, 2016, pp. 86–107. doi: 10.1007/978- 3-319-44902-9_7.

[32]. L. P. Binamungu, S. M. Embury, and N. Konstantinou, “Characterising the Quality of Behaviour Driven Development Specifications,” in Agile Processes in Software Engineering and Extreme Programming. XP 2020. Lecture Notes in Business Information Processing, Springer, Cham, 2020, pp. 87–102. doi: 10.1007/978-3-030-49392-9_6.

[33]. T. R. Silva, M. Winckler, and H. Trætteberg, "Ensuring the Consistency between User Requirements and Task Models: A Behavior-Based Automated Approach," Proc. ACM Hum.-Comput. Interact., vol. 4, no. EICS, pp. 1–32, 2020, doi: 10.1145/3394979.

[34]. L. P. Binamungu, S. M. Embury, and N. Konstantinou, "Maintaining behaviour driven development specifications: Challenges and opportunities," in 25th IEEE International Conference on Software Analysis, Evolution and Reengineering, SANER 2018 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 175–184. doi: 10.1109/SANER.2018.8330207.

[35]. T. N. Kudo, R. de F. Bulcão-Neto, V. V. G. Neto, and A. M. R. Vincenzi, "Aligning requirements and testing through metamodeling and patterns: design and evaluation," Requir Eng, vol. 28, no. 1, pp. 97–115, Mar. 2023, doi: 10.1007/s00766-022-00377-5.

[36]. Y. Wang, D. R. Degutis, and S. Wagner, "Speed up BDD for safety verification in agile development: A partially replicated controlled experiment," in ACM International Conference Proceeding Series, Association for Computing Machinery, 2018. doi: 10.1145/3234152.3234181.

[37]. G. Oliveira, S. Marczak, and C. Moralles, "How to evaluate BDD scenarios' quality?" in Proceedings of the XXXIII Brazilian Symposium on Software Engineering, Association for Computing Machinery, 2019, pp. 481–490. doi: 10.1145/3350768.3351301.

[38]. N. Patkar, A. Chis, N. Stulova, and O. Nierstrasz, "Interactive Behavior-driven Development: A Low-code Perspective," in 24th International Conference on Model-Driven Engineering Languages and Systems, MODELS-C 2021, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 128–137. doi: 10.1109/MODELS-C53483.2021.00024.

[39]. T. Zameni, P. Van Den Bos, J. Tretmans, J. Foederer, and A. Rensink, "From BDD Scenarios to Test Case Generation," in Proceedings - 2023 IEEE 16th International Conference on Software Testing, Verification and Validation Workshops, ICSTW 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 36–44. doi: 10.1109/ICSTW58534.2023.00019.

[40]. S. Rodriguez, J. Thangarajah, M. Winikoff, and M. 2023 Winikoff, "A Behaviour-Driven Approach for Testing Requirements via User and System Stories in Agent Systems. In Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems," 2023.

[41]. S. Heng, K. Tsilionis, and Y. Wautelet, "Building User Stories and Behavior Driven Development Scenarios with a Strict Set of Concepts: Ontology, Benefits and Primary Validation," in Proceedings of the ACM Symposium on Applied Computing, Association for Computing Machinery, Mar. 2023, pp. 1422–1429. doi: 10.1145/3555776.3577696.

[42]. J. Saraiva and S. Soares, "Adoption of the LGPD Inventory in the User Stories and BDD Scenarios Creation," in ACM International Conference Proceeding Series, Association for Computing Machinery, Sep. 2023, pp. 416–421. doi: 10.1145/3613372.3613375.

[43]. J. A. Valdivia, A. Lora-González, X. Limón, K. Cortes-Verdin, and J. O. Ocharán-Hernández, "Patterns Related to Microservice Architecture: A Multivocal Literature Review," Programming and Computer Software, vol. 46, no. 8, pp. 594-608, Dec. 2020, doi: 10.1134/S0361768820080253.

[44]. R. Moguel-Sánchez, C. S. Sergio Martínez-Palacios, J., O. Ocharán-Hernández, X. Limón, and A. J. Sánchez-García, "Bots in Software Development: A Systematic Literature Review and Thematic Analysis," Programming and Computer Software, vol. 49, no. 8, pp. 712-734, Dec. 2023, doi: 10.1134/S0361768823080145.

Информация об авторах / Information about authors

Виктор Мануэль АРРЕДОНДО-РЕЙЕС – студент бакалавриата по направлению программной инженерии факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных интересов: проектирование, разработка и тестирование программного обеспечения, инженерия требований к программному обеспечению.

Víctor Manuel ARREDONDO-REYES, Undergraduate Student in Software Engineering, Facultad de Estadística e Informática, Universidad Veracruzana, Mexico (School of Statistics and Informatics, University of Veracruz). Research interests: software design, requirements engineering, software development, and software testing.

Саул ДОМИНГЕС-ИСИДРО имеет степень PhD по искусственному интеллекту, доцент факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных

интересов: распределенные системы, разработка программного обеспечения, вычислительный интеллект, машинное обучение.

Saúl DOMÍNGUEZ-ISIDRO – PhD in artificial intelligence, associate professor at Facultad de Estadística e Informática, Universidad Veracruzana, Mexico (School of Statistics and Informatics, University of Veracruz). Research interests: distributed systems, software development, computational intelligence, and machine learning.

Ангел Хуан САНЧЕС-ГАРСИЯ имеет степень PhD по искусственному интеллекту, доцент факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных интересов: программометрия, машинное обучение, прогнозирование затрат, эволюционные вычисления.

Ángel Juan SÁNCHEZ-GARCÍA – PhD in artificial intelligence, associate professor at Facultad de Estadística e Informática, Universidad Veracruzana, Mexico (School of Statistics and Informatics, University of Veracruz). Research interests: software measurement, machine learning, effort prediction, and evolutionary computation.

Хорхе Октавио ОЧАРАН-ЭРНАНДЕС имеет степень PhD по искусственному интеллекту, доцент факультета статистики и информатики Университета Веракруса (Мексика). Сфера научных интересов: архитектура программного обеспечения, инженерия требований к программному обеспечению, программная инженерия, проектирование прикладных интерфейсов.

Jorge Octavio OCHARÁN-HERNÁNDEZ – PhD in Computer Science, Associate Professor at the Facultad de Estadística e Informática, Universidad Veracruzana, Mexico (School of Statistics and Informatics, University of Veracruz). Research interests: software architecture, requirements engineering, software engineering, and API design.

DOI: 10.15514/ISPRAS-2024-36(6)-10

Secure and Efficient Data Model for Public Lighting in México with AMI/IoT: Implementing LZ4 Compression, IPFS, and Blockchain

[1] *R. García Reyes, ORCID: 0009-0002-7395-9362 <d21ce036@cenidet.tecnm.mx>*
[1] *J. Ortiz-Hernandez, ORCID: 0000-0002-6481-1537 <javier.oh@cenidet.tecnm.mx>*
[2] *R. Mijarez, ORCID: 0000-0001-8369-4092 <r.mijarez.c@gmail.com>*
[3] *J. A. Hernández-Aguilar, ORCID: 0000-0002-5184-0005 <jose_hernandez@uaem.mx>*
[1] *Y. Hernández, ORCID: 0000-0002-8842-0899 <yasmin.hp@cenidet.tecnm.mx>*

[1] *National Technological Institute of Mexico, National Center for Research and Technological Development, Cuernavaca, México.*
[2] *National Institute of Electricity and Clean Energy, Cuernavaca, México.*
[3] *Autonomous University of the State of Morelos, Cuernavaca, México*.

Abstract. The advent of digitalization and Internet of things (IoT) technologies brings new challenges to the management of electric metering systems. Integrating institutional energy billing systems with government Ambient Intelligence (AMI) systems is essential for effective management. Blockchain technology is proposed to maintain data integrity through automated energy readings. This study introduces an innovative model designed to enhance public lighting in Mexico by integrating AMI and IoT, and employing LZ4 and IPFS for data compression. This approach aims to optimize the handling of large data volumes, resulting in improved data efficiency, enhanced security, cost reductions, and better energy resource management.

Keywords: blockchain; Internet of things (IoT); advanced metering infrastructure (AMI); inter planetary file system (IPFS); LZ4; public lighting.

For citation: García Reyes R., Ortiz-Hernandez J., Mijarez R., Hernández-Aguilar J. A., Hernández Y. Secure and Efficient Data Model for Public Lighting in México with AMI/IoT: Implementing LZ4 Compression, IPFS, and Blockchain. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 179-194. DOI: 10.15514/ISPRAS-2024-36(6)-10.

Acknowledgements. This research was supported by the National Center for Research and Technological Development (CENIDET) through the CONACYT doctoral scholarship. RGR acknowledges the financial support of CONACYT, Mexico.

García-Reyes R., Ortiz-Hernandez J., Mijarez R., Hernández-Aguilar J. A., Hernández Y. Secure and Efficient Data Model for Public Lighting in México with AMI/IoT: Implementing Lz4 Compression, IPFS, and Blockchain. *Trudy ISP RAN/Proc. ISP RAS*, vol. 36, issue 6, 2024. pp. 179-194.

Безопасная и эффективная модель данных для электрического освещения общественного пространства в Мексике с AMI/IoT: реализация сжатия LZ4, IPFS и блокчейна

[1] *Р. Гарсия Рейес, ORCID: 0009-0002-7395-9362 <d21ce036@cenidet.tecnm.mx>*
[1] *Х. Ортис-Эрнандес, ORCID: 0000-0002-6481-1537 <javier.oh@cenidet.tecnm.mx>*
[2] *Р. Михарес, ORCID: 0000-0001-8369-4092 <r.mijarez.c@gmail.com>*
[3] *Х. А. Эрнандес-Агилар, ORCID: 0000-0002-5184-0005 <jose_hernandez@uaem.mx>*
[1] *Я. Эрнандес, ORCID: 0000-0002-8842-0899 <yasmin.hp@cenidet.tecnm.mx>*

[1] *Национальный центр технологических исследований и разработок Мексики, Куэрнавака, Мексика.*
[2] *Национальный институт электроэнергии и чистой энергетики, Куэрнавака, Морелос, Мексика.*
[3] *Автономный университет штата Морелос, Куэрнавака, Мексика.*

Аннотация. Появление цифровизации и технологий Интернета вещей (IoT) ставит новые задачи по управлению системами учета электроэнергии. Интеграция институциональных систем выставления счетов за электроэнергию с государственными измерительными инфраструктурными системами (AMI) имеет важное значение для эффективного управления. Для обеспечения целостности данных при автоматизации снятия показаний электроэнергии предлагается использовать технологию распределенного реестра (блокчейн). Настоящая работа описывает инновационную модель, предназначенную для улучшения освещения общественного пространства в Мексике путем интеграции систем AMI и IoT, а также использования алгоритмов LZ4 и IPFS для сжатия данных. Выбранный подход направлен на оптимизацию обработки больших объемов данных, его использование приводит к повышению эффективности данных, повышению безопасности, снижению затрат и улучшению управления энергоресурсами.

Ключевые слова: блокчейн; интернет вещей (IoT); окружающий интеллект (AMI); межпланетная файловая система (IPFS); алгоритм сжатия LZ4; электрическое освещение общественного пространства.

Для цитирования: Гарсия-Рейес Р., Ортис-Эрнандес Х., Михарес Р., Эрнандес-Агилар Х. А., Эрнандес Я. Безопасная и эффективная модель данных для электрического освещения общественного пространства в Мексике с AMI/IoT: реализация сжатия LZ4, IPFS и блокчейна. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 179–194 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–10.

Благодарности. Выполнение исследования поддерживалось Национальным центром исследований и технологического развития (CENIDET) через докторскую стипендию CONACYT. Первый автор выражает благодарность за финансовую поддержку CONACYT, Мексика.

1. Introduction

Blockchain technology, through its distributed ledger technology (DLT), offers enhanced security and transparency for energy transactions, particularly in peer-to-peer (P2P) trading. Despite facing challenges in technology, economy, society, and regulation, its potential to revolutionize microgrid management and automate billing processes is significant [1]. In Mexico's advanced metering infrastructure (AMI) for public lighting, has improved real-time energy management and billing, but discrepancies due to system differences and manual interventions can lead to billing inaccuracies and loss of consumer trust [2]. Blockchain could address these issues by providing immutable and transparent records, acting as an integration bridge between institutional and proprietary AMI systems in public lighting. Beyond finance, real estate, insurance, and supply chain sectors, blockchain demonstrates its value in solving complex energy management problems.

2. Definition of the problem

In Mexico, the use of IoT electric meters is limited and involves addressing security, privacy, and reliability concerns in critical environments, with blockchain, IPFS (Inter Planetary File System) [3], and LZ4 [4] proposed for secure data handling and distribution.

The research aims to develop a model for securely managing AMI/IoT meter data in public lighting systems, which handle thousands of devices per customer and seek improved billing processes. Integrating blockchain in public lighting presents challenges, including the need for a robust infrastructure to support secure P2P transactions [5-6], addressing increased bandwidth and latency issues and handling high data volumes efficiently [7].

2.1 Objectives

1. Propose a Model: Develop a data integration model to secure and protect information from IoT devices in utility power grids, such as street lighting, handling thousands of AMI/IoT devices.
2. Implement Protection Mechanisms: Use IPFS and Blockchain for data integrity and authenticity, with LZ4 for additional compression.
3. Evaluate Effectiveness: Assess the system's security, performance, and scalability in IoT meter power grid environments.

2.2 Hypothesis

The combination of Blockchain, IPFS, and LZ4 aims to secure and optimize IoT data integration in utility power grids, like street lighting. Unlike existing models, which focus on domestic and industrial uses, this solution targets IoT meters for single customers, such as street lighting. It will enhance data efficiency and security, lower costs, and improve energy management while ensuring data integrity in dynamic environments with numerous AMI/IoT devices.

2.3 Contribution

This work's main contribution is a new model for addressing security and privacy challenges in AMI/IoT data integration for utility power grids. By combining IPFS and Blockchain with LZ4 compression, the model offers a robust solution that enhances the reliability and security of power grids, leading to more efficient and secure services for users.

3. Related work

In recent years, AMI has transformed energy measurement and management, including in street lighting [8]. However, the integration of various systems has created operational challenges [9]. Blockchain technology offers a potential solution for data integrity, decentralization, and transparency [10], with ongoing studies examining its use in the energy sector [11].

The rapid evolution of smart grids and smart meters has raised significant data security and communication challenges. These meters, crucial for AMI, are exposed to cyber-attacks through public channels, such as data theft and unauthorized access [12]. Islam's study [12] proposes using blockchain to secure AMI communication, addressing issues with traditional cryptographic solutions that rely on trusted third parties. Their blockchain-based approach, using smart contracts and PBFT, aims to reduce costs and enhance security in AMI systems.

The Decentralized Autonomous Area (D3A) model uses blockchain and smart contracts to enhance and decentralize the power grid, improving the integration of renewable resources and system resilience by fostering transparency and competition [13]. It also securely integrates IoT data, addressing power grid security challenges.

The article "Decentralized Energy Networks Based on Blockchain" [14] highlights the role of blockchain in the decentralization of the energy sector, including local renewable energy communities and P2P markets. This context is relevant for understanding how IPFS and blockchain can facilitate secure transitions in energy systems.

The TWACS (Two Way Automatic Communication System) telemetering system [15], used by CFE in Mexico [16], faced security issues due to its centralized structure and lack of blockchain and IPFS integration. While it managed large volumes of data efficiently, its vulnerability to file modifications highlighted the need for improved security and decentralized data management.

3.1. Blockchain and IPFS on AMI systems

Blockchain technology, introduced by Satoshi Nakamoto in 2008 and launched in 2009, is a decentralized digital ledger that stores encrypted information, ensuring data integrity and preventing tampering [17]. Each block is linked to previous ones, and replication across multiple nodes guarantees data originality. This technology holds significant promise for transforming AMI by offering a secure, transparent foundation for operations [18].

Originally associated with cryptocurrencies, blockchain's potential for AMI systems has become evident. There are two main types of blockchains: public (permissionless) and private (permissioned). Public blockchains are open to all, while private ones restrict participation to selected nodes, balancing transparency with privacy for AMI systems [17-18]. Smart contracts, which automate processes like reading and billing, reduce costs by eliminating intermediaries [19].

The Hyperledger Fabric (HLF) platform by IBM is a notable example of blockchain technology applied to IoT/AMI systems. It features an "execute-order" model, optimizing performance and confidentiality, crucial for dynamic AMI environments [20].

The article "Blockchain-Based Applications for Smart Grids: An Umbrella Review" [21] provides a comprehensive review of blockchain applications in smart grids, synthesizing findings from various studies to address technical inaccuracies and highlight blockchain's transformative potential for modernizing power grids.

The Umbrella review model is presented in Fig. 1 with the most representative elements of the work.

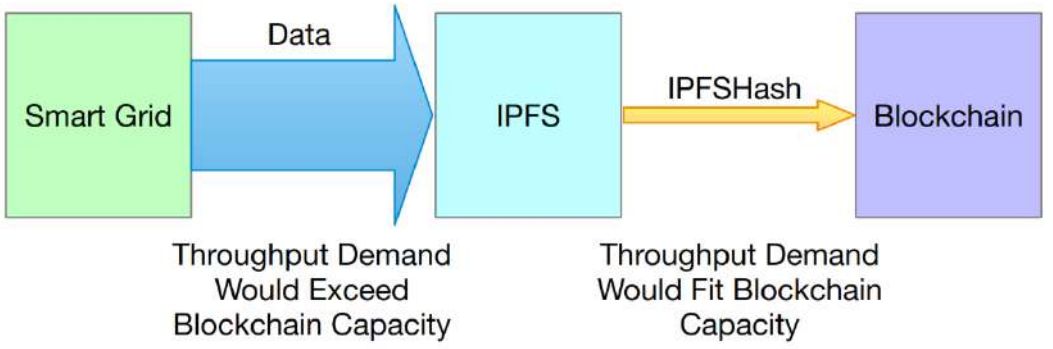


Fig. 1. Data flow of the Blockchain-Based Applications for Smart Grids: An Umbrella Review model [20].

Méndez [22] suggests developing dynamic smart communities in Mexico to promote energy savings through smart interfaces. The study notes that the residential sector consumed nearly 39% of electricity in 2021 and emphasizes the impact of user habits on consumption. It advocates for using gamification and artificial neural networks (ANNs) to improve user interaction and design more personalized, energy-efficient systems.

Electricity losses, including theft and billing irregularities, pose a major challenge globally, especially in developing countries [23]. Innovative approaches to consumption, management, and billing are needed.

Current systems, even with measures like prepaid metering, have limited impact on reducing losses [23]. Strengthening regulations and security measures, along with improving customer relationships, could be more effective. The proposed model addresses environmental concerns by using less energy-intensive consensus mechanisms like Proof-of-Stake (PoS) or Proof-of-Authority (PoA), aligning with sustainability goals [24].

Blockchain's ability to facilitate secure, direct transactions between prosumers without intermediaries enhances privacy and trust in energy trading [25]. Integrating blockchain with AMI systems can reduce billing irregularities and fraud, automate processes, and build trust between customers and utilities.

The article "Managing the computational load in a Blockchain-based multilevel IoT network" [26] discusses optimizing computational load by dividing the IoT network into layers, blending centralized IoT with blockchain distribution to avoid disrupting existing applications. This work helps understand blockchain integration in IoT networks and informs the design of similar models for utility power grids.

The paper "A Method for Protecting Private Data in IPFS" describes a method combining blockchain and enhanced IPFS to protect private data by managing permissions and file organization [27]. This approach could enhance data security and privacy in the developed model for IoT and street lighting systems.

Another model, the "Blockchain-based multilayer model," divides data into four layers to improve public information integrity and security, using different consensus algorithms for each layer to ensure decentralization and transaction speed [28]. This model aims to provide secure and transparent digital government services.

Fig. 2 shows the general diagram of the proposed model, with its most important and representative elements, as well as the stakeholders involved in the processes.

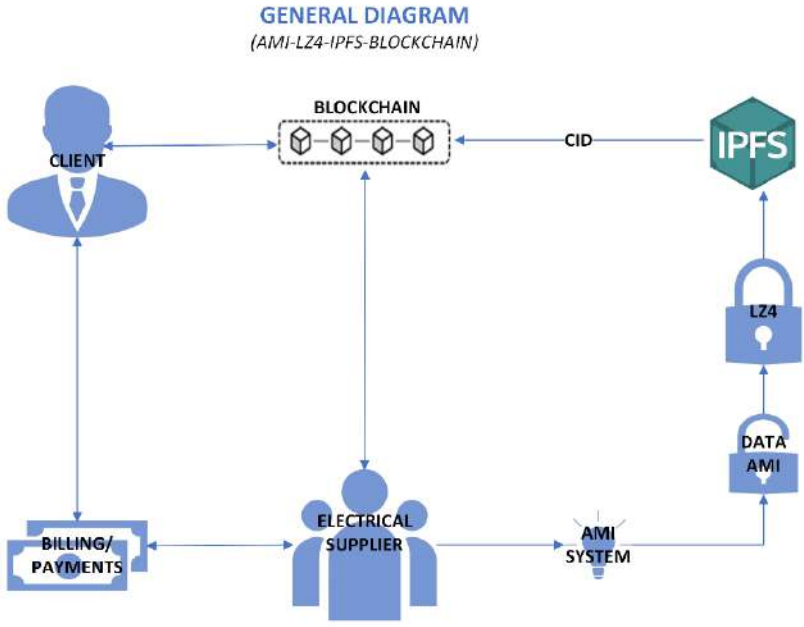


Fig. 2. General diagram, own elaboration.

3.2 Methodology

- Problem Definition:
- Create a model for secure AMI/IoT data integration in Mexico's power grids using blockchain, IPFS, and LZ4.
- Define data flow and security requirements.
- Specification of Requirements:
- Set objectives for data security and access.
- Define software-level requirements.

1. Research Question Posing:
   - Formulate questions on technologies and methods.
   - Explore development approaches.
2. Review of Methods:
   - Review existing AMI/IoT data integration projects.
3. Identification of Relevance:
   - Assess the proposed solution's relevance and effectiveness.
4. Design and Development:
   - Develop a blockchain and IPFS solution with strong security and consensus protocols.
5. Expert Feedback:
   - Gather and incorporate expert feedback.

4. Model development

The model aimed to improve street lighting under APBT and APMT tariffs [29] using AMI and IoT technologies. These tariffs (see Table 1), previously 5 and 5A, were reformed in Mexico in 2014 and split into basic supply and distribution in 2016 [30-31].

Current AMI/IoT models focus on domestic and commercial sectors, where each meter serves an individual customer. In Mexico, municipalities face a unique challenge with a single account managing thousands of devices. This case study highlights the need for scalable, interoperable, and secure AMI/IoT solutions to handle the complexities of large-scale public lighting.

Table 1. Classification of electricity tariffs in Mexico by customer, energy demand, and associated AMI/IoT meter model

Tariff Category	Description	Previous Tariff	Commercial Model
Domestic	Domestic Low Voltage	1, 1A-F	1 customer - 1 meter
PDBT	Small Demand Low Voltage	2, 6	1 customer - 1 meter
GDBT	Large Demand Low Voltage	3, 6	1 customer - 1 meter
RABT	Agricultural Irrigation Low Voltage	9	1 customer - 1 meter
APBT	Public Lighting Low Voltage	5	1 customer - Thousands of meters
APMT	Public Lighting Medium Voltage	5A	1 customer - Thousands of meters
GDMTH	Large Demand Hourly Medium Voltage	HM, HMC, 6	1 customer - 1 meter
GDMTO	Large Demand Ordinary Medium Voltage	OM, 6	1 customer - 1 meter
RAMT	Agricultural Irrigation Medium Voltage	9M	1 customer - 1 meter
DIST	Industrial Sub Transmission	HS, HSL	1 customer - 1 meter
DIT	Industrial Transmission	HT, HTL	1 customer - 1 meter

From a technical standpoint, the AMI/IoT system designed by INEEL in Mexico [32] is used within the Enabling Technologies Division's Control, Electronics, and Communications department. This system utilizes RF communication and ARM architecture, as shown in Fig. 3.
Three Meters generates their files using the JSON format, with the structure depicted in Fig. 4.

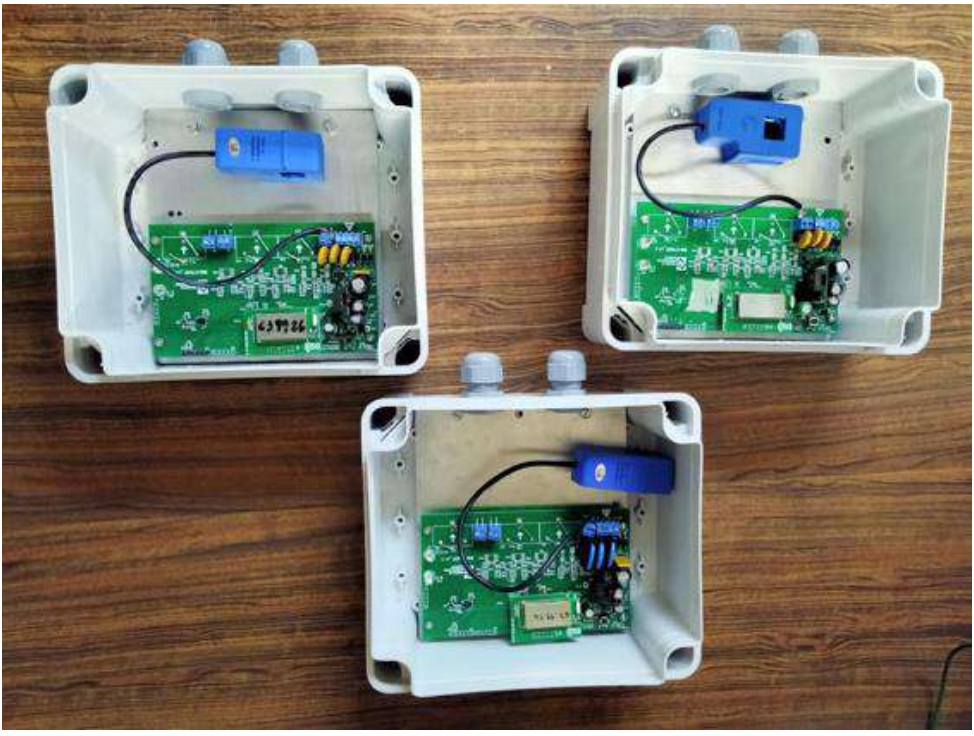


Fig. 3. AMI/IoT meter, property of INEE.


```
{    "Id": "7538PH",    "Consumo": 37354,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0000CU",    "Consumo": 45096,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0000NC",    "Consumo": 70776,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "00010C",    "Consumo": 58442,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "00010G",    "Consumo": 93780,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "00010J",    "Consumo": 85023,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "00010Q",    "Consumo": 83475,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "00010T",    "Consumo": 71758,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "00010X",    "Consumo": 10801,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001AL",    "Consumo": 56274,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001BD",    "Consumo": 27011,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001CH",    "Consumo": 33659,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001CI",    "Consumo": 52655,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001DP",    "Consumo": 51635,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001FL",    "Consumo": 54779,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001GH",    "Consumo": 38015,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001HV",    "Consumo": 62393,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001ID",    "Consumo": 70423,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001LB",    "Consumo": 25809,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001OD",    "Consumo": 57701,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001PH",    "Consumo": 64308,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001QK",    "Consumo": 24803,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001TH",    "Consumo": 16732,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001VY",    "Consumo": 43254,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0001XU",    "Consumo": 95622,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0002BV",    "Consumo": 15112,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0002BZ",    "Consumo": 71163,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
{    "Id": "0002DI",    "Consumo": 85645,    "Fecha": "45105",    "TipoMedidor": 1,    "Notas": ""  },
```


Fig 4. AMI/IoT file structure, property of INEEL.

The data format used is JSON, but XML or plain text are also common in the AMI industry. Meters include an ID, date, and type. To simulate a large-scale scenario with thousands of AMI/IoT meters for street lighting, synthetic data subsets were created, recording only consumption in kWh under

low-voltage conditions. Data size is further reduced using LZ4 compression before being transferred to IPFS, adding an extra layer of security. The compressed files can be previewed in a text editor, as shown in Fig. 5.

The LZ4 algorithm was chosen for its excellent compatibility with the data, as well as its efficient transfer, compression, and decompression capabilities. Its ease of integration into Visual Studio Community 2022, due to the availability of libraries and resources, further supported its selection. Fig. 6 illustrates LZ4's performance in compression, transfer, and decompression.

After compressing AMI/IoT files with LZ4, they are transferred to the IPFS, a distributed file system that decentralizes file storage and distribution across multiple nodes instead of relying on a single server.

To transfer files to IPFS, follow these steps:

- Initialize IPFS Node: Install IPFS software on the device and run the initialization command.
- Add Files: Upload LZ4-compressed AMI files to IPFS using the command, which assigns a unique CID to each file.
- Distribute Files: IPFS automatically propagates the files to other nodes in the network for efficiency and redundancy.
- Access Files: Retrieve files using their CID, which provides a unique address for access from any IPFS node.

Fig. 7 shows files loaded into IPFS with LZ4 compression applied. Fig. 8 illustrates the generation of a CID (Content Identifier) in IPFS, which enables easy retrieval from anywhere in the distributed network. This process also applies to files with synthetic data.

Using IPFS for storing and distributing LZ4-compressed AMI files offers an efficient, cost-effective, and resilient solution:

- Reduced Bandwidth Usage: IPFS's decentralized network distributes content across multiple nodes, decreasing load on any single server and reducing bandwidth usage.
- Lower Storage Cost: IPFS utilizes storage across various nodes, minimizing the need for a large centralized server and cutting storage costs.
- Censorship Resistance: IPFS's decentralized nature prevents censorship and single points of failure, ensuring long-term availability and access by all relevant parties.

Fig. 5. File structure with the LZ4 [4] algorithm applied.

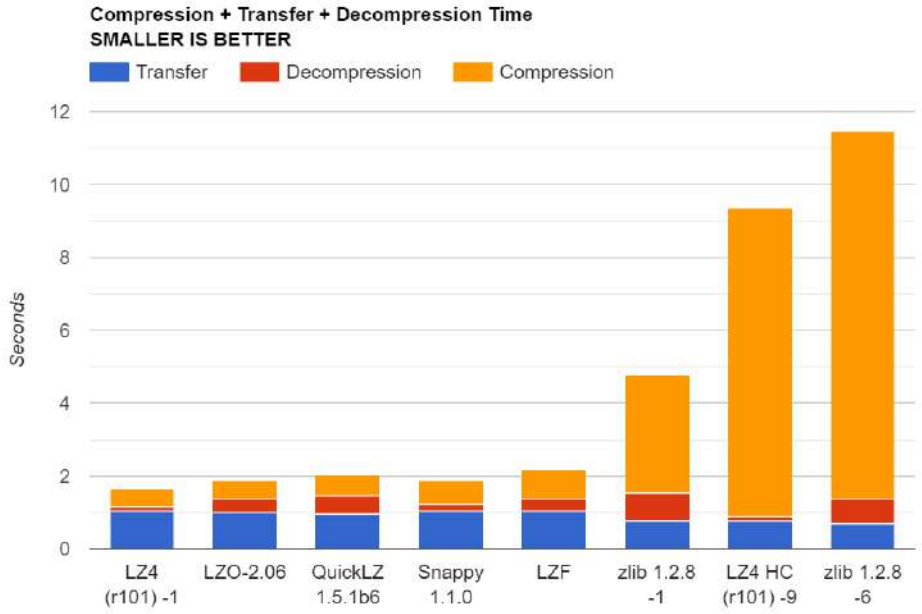


Fig. 6. Comparison of compression, transfer, and decompression levels with working times of the LZ4 algorithm [4].

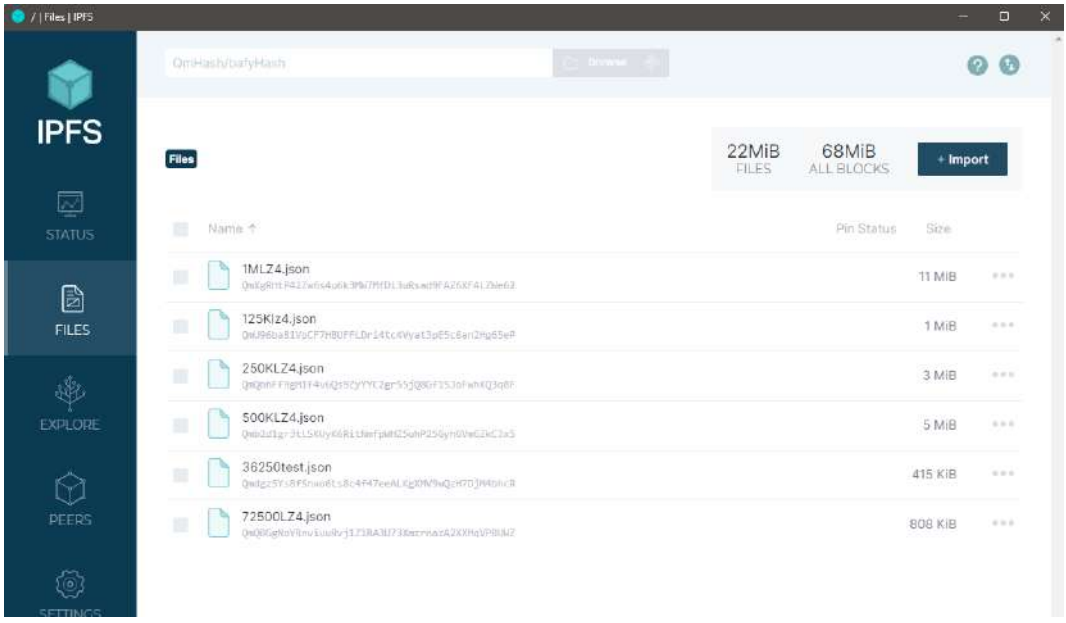


Fig. 7. View of files loaded in the IPFS already with LZ4 compression applied.

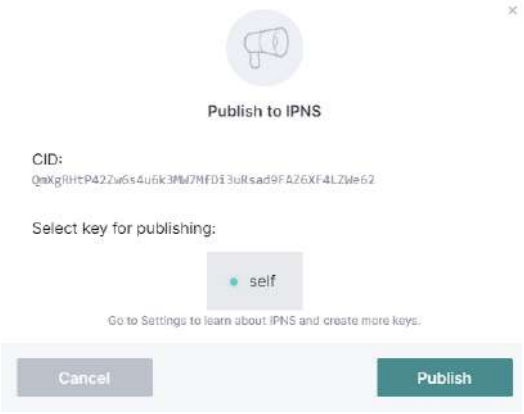


Fig. 8. CID (Content Identifier) generation in IPFS.

Table 2 provides comprehensive data on the optimization process, presenting file sizes before and after compression, optimization percentages, and associated costs in Ether (ETH), USD, and MXN. The table highlights the substantial reductions in both file size and costs, demonstrating improvements in system efficiency through the use of LZ4 and JWT compression.

Table 2. Final results.

Readings per meter	File Size (kB)	File Size after LZ4 and JWT (kB)	% Optimization	Cost in Ether (ETH)	Cost (USD)	Estimated Gas Required	Total Cost to Upload to Blockchain USD (original size)	Total Cost to Upload to Blockchain USD (optimized)	Cost in MXN Optimized	Cost in MXN Unoptimized
1M	106446	11172	89.5	0.0002	0.31	200000	$ 31.58	$ 0.31	$ 5.56	$ 99.71
500K	50787	5564	89.04	0.00015	0.23	150000	$ 15.79	$ 0.23	$ 4.12	$ 73.98
250K	25391	2763	89.12	0.0001	0.16	100000	$ 7.89	$ 0.16	$ 2.87	$ 51.46
125K	12696	1362	89.27	0.00008	0.12	80000	$ 3.95	$ 0.12	$ 2.15	$ 38.60
72.5K	7364	809	89.01	0.00006	0.09	60000	$ 2.30	$ 0.09	$ 1.61	$ 28.95
36.2K	3682	416	88.7	0.00004	0.06	40000	$ 1.15	$ 0.06	$ 1.08	$ 19.30

The model effectively reduces file size using LZ4 and JWT, while IPFS and blockchain ensure efficient data management and integrity.

Results indicate a significant reduction in storage and transaction costs through data compression. The combination of LZ4 and JWT with IPFS and blockchain enhances data management by offering efficiency, robustness, and cost-effectiveness.

The proposed model, integrating advanced compression and distributed storage, surpasses traditional AMI systems. Unlike centralized systems, this model leverages blockchain's immutable security and IPFS's distributed nature, which together prevent data manipulation and increase resilience.

Though blockchain storage incurs transaction costs, the reduction in file sizes and the use of IPFS counterbalance this expense, leading to long-term savings and improved data management.

Fig. 9 illustrates the AMI-LZ4-IPFS-HASH (ALIH) model, comprising four layers:

1. AMI/IoT Data Layer: Generates and loads data from the Smart Grid.
2. LZ4 Compression Layer: Applies compression to the data.

3. IPFS Layer: Stores and distributes the compressed data.
4. Hash-CID Layer: Manages file identification and retrieval via CID.

Table 3 compares a traditional AMI system with the enhanced model using IPFS and blockchain. The enhanced model provides superior data protection, integrity, and resilience through decentralized storage and immutable verification, resulting in reduced costs and improved security over the traditional centralized approach.

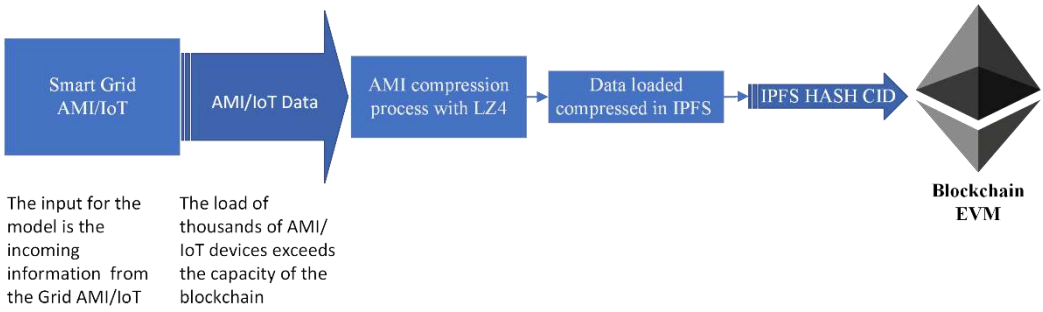


Fig. 9. ALIH Model Diagram of Blocks.

Table 3. Traditional AMI System versus Proposed Model.

Aspect	Traditional AMI System	Enhanced Model with IPFS, LZ4 and Blockchain
Encryption Standards	AES-128 and AES-256 for encryption in transit and at rest.	LZ4 for compression, JWT for authentication, Blockchain for integrity.
Data Protection	Centralized protection, depends on server and network security.	Decentralized protection, immutable verification via blockchain.
Data Authenticity	Based on internal verification mechanisms, vulnerable to attacks.	Ensured by JWT and immutable records in the blockchain.
Data Integrity	Controlled by audit logs and access controls, susceptible to alterations.	Ensured by cryptographic hashes in blockchain (SHA-256).
Resilience	Depending on server and network redundancy, may fail under DDoS attacks.	High resilience through distributed storage in IPFS and immutable copies in blockchain.
Data Availability	May be affected by server or network failures.	High availability ensured by distribution across IPFS nodes and replication in blockchain.
Storage Costs	High due to centralized infrastructure and server maintenance.	Reduced through compression and distributed storage in IPFS; gas costs on blockchain offset by reduced file size.
Security against Tampering	Vulnerable to internal manipulations and attacks if central server is compromised.	Robust protection through blockchain immutability and cryptographic verification.
Transaction and Gas Costs	Not applicable, costs are mainly related to central infrastructure.	Costs related to blockchain gas; significant optimization due to reduced file size.

5. Conclusions

The proposed ALIH model has proven effective in terms of both security and economic efficiency, as supported by the data in Table 2 and the completion of planned activities. Key aspects include:

5.1 Benefits

Security: LZ4 compression enhances security by reducing file size, making data tampering harder. IPFS ensures integrity and decentralization, reducing the risk of data alteration or loss.

Economic Efficiency: LZ4 compression significantly cuts storage and transfer costs on the blockchain, improving cost management.

Practical Implementation: The model successfully executed all planned tasks, including JSON file generation, LZ4 compression, and IPFS uploading, validating its effectiveness in security and cost efficiency.

5.2 Limitations

Blockchain Costs: Gas fees can still be a limiting factor, especially for large-scale implementations.

Scalability: Further exploration is needed to ensure consistent performance with very high data volumes or varying network conditions.

5.3 Future Research

Performance Optimization: Focus on enhancing compression and blockchain integration to further reduce costs and boost speed.

Scalability Studies: Investigate the model's performance across different network environments and data volumes.

Security Enhancements: Explore advanced encryption and multi-factor authentication to strengthen security.

In summary, the ALIH model addresses security and economic challenges in AMI data management, with successful implementation and cost reductions validating its potential. Future research should refine and expand the model's capabilities for broader application.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

[1]. Evens, M., Ercoli, P., & Arteconi, A., "Blockchain-Enabled Microgrids: Toward Peer-to-Peer Energy Trading and Flexible Demand Management", Energies, vol. 16, no. 16, p. 6741, 2023. [Online]. Available: https://doi.org/10.3390/en16186741.

[2]. CONUEE, "Proyecto piloto de tele gestión en alumbrado público", Dirección de Fomento, Difusión e Innovación, Comisión Nacional para el Uso Eficiente de la Energía, January 2023. [Online]. Available: https://www.gob.mx/conuee/articulos/publica-la-conuee-el-informe-del-proyecto-piloto-de-telegestion?idiom=es.

[3]. IPFS "For Developers: Navigating IPFS", IPFS.tech. [Online]. Disponible en: https://ipfs.tech/developers/. [Access: 4 may 2024].

[4]. LZ4 Block Format Description. [Online]. Disponible en: https://lz4.org/ [Access: 4 may 2024]

[5]. Choobineh, M., Paaso, A., Arabnya, A., Sohrabi, B., & Khodaei, A. (2023). Blockchain technology in energy systems: A state-of-the-art review. IET Blockchain, 3(1), 35–59. https://doi.org/10.1049/blc2.12020.

[6]. Andoni, M., Robu, V., Flynn, D., Abram, S., Geach, D., Jenkins, D., et al.: Blockchain technology in the energy sector: A systematic review of challenges and opportunities. Renewable Sustainable Energy Rev. 100,143–174 (2019).

[7]. Meeuw, A., Schopfer, S., Wortmann, F.: Experimental bandwidth benchmarking for P2P markets in blockchain managed microgrids. Energy Procedia. 159, 370–375 (2019).

[8]. Gungor, V. C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., & Hancke, G. P. (2013). "Smart Grid and Smart Homes: Key Players and Pilot Projects", IEEE Industrial Electronics Magazine, 7(4), 18- 34.

[9]. Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of Things for Smart Cities. IEEE Internet of Things Journal, 1(1), 22-32.

[10]. Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction. Princeton: Princeton University Press.

[11]. Mengelkamp, E., Notheisen, B., Beer, C., Dauer, D., & Weinhardt, C. (2018). A blockchain-based smart grid: towards sustainable local energy markets. Computer Science - Research and Development, 33(1-2), 207-214.

[12]. Islam, N., Rahman, M. S., Mahmud, I., Sifat, M. N. A., & Cho, Y.-Z. (2022). A Blockchain-Enabled Distributed Advanced Metering Infrastructure Secure Communication (BC-AMI). Appl. Sci., 12(14), 7274. https://doi.org/10.3390/app12147274.

[13]. Bronski, P., Creyts, J., Gao, S., Hambridge, S., Hartnett, S., Hesse, E., Morris, J., Nanavatty, R., & Pennington, N. (2018). "The Decentralized Autonomous Area Agent (D3A) Market Model: An Implementation Framework for Transactive Energy Based on Blockchain for the 21st Century Grid", Energy Web Foundation. Retrieved from https://energyweb.org/D3A.

[14]. Pichler, M., Meisel, M., Goranovic, A., Leonhartsberger, K., Vallant, H., Lettner, G., Marksteiner, S., Chasparis, G., & Bieser, H. (2019). "Decentralized Energy Networks Based on Blockchain: Background, Overview and Concept Discussion" in W. Abramowicz & A. Paschke (Eds.), BIS 2018 Workshops (LNBIP 339, pp. 244–257). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-04849- 5_22.

[15]. Aclara, "TWACS Net Server with TWACS OC Software", Hubbell, 2008. [Online]. Available: https://www.hubbell.com/aclara/en/solutions/twacs. [Accessed: May 16, 2024].

[16]. G. Reyes, "Enterprise Management Systems for Database Administration Focused on Invoicing Processes", Master's thesis, Facultad de Contaduría Administración e Informática, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, Noviembre 2014.

[17]. Aggarwal, S., Chaudhary, R., Aujla, G.S., Kumar, N., Choo, K.K.R., & Zomaya, A.Y. (2019). Blockchain for smart communities: Applications, challenges and opportunities. J. Netw. Comput. Appl., 144, 13–48.

[18]. Liu, Z., Luong, N.C., Wang, W., Niyato, D., Wang, P., Liang, Y.-C., & Kim, D.I. (2019). A survey on applications of game theory in blockchain. arXiv, arXiv:1902.10865.

[19]. Wang, S, Ouyang, L., Yuan, Y., Ni, X., Han, X., & Wang, F.Y. (2019). Blockchain-enabled smart contracts: Architecture, applications, and future trends. IEEE Trans. Syst. Man Cybern. Syst., 49, 2266–2277.

[20]. Graf, M., Küsters, R., & Rausch, D. (2020). Accountability in a permissioned blockchain: Formal analysis of hyperledger fabric. In Proceedings of the 2020 IEEE European Symposium on Security and Privacy (EuroS&P), Genoa, Italy, 7–11 September 2020.

[21]. Zhao, W., Qi, Q., Zhou, J., & Luo, X. (2023). Blockchain-Based Applications for Smart Grids: An Umbrella Review. Energies, 16(6147). https://doi.org/10.3390/en16176147.

[22]. Méndez, J. I., Medina, A., Ponce, P., Peffer, T., Meier, A., & Molina, A. (2022). Evolving Gamified Smart Communities in Mexico to Save Energy in Communities through Intelligent Interfaces. Energies, 15(5553). https://doi.org/10.3390/en15155553.

[23]. Carr, D., & Thomson, M. (2022). Non-Technical Electricity Losses. Energies, 15(6), 2218. https://doi.org/10.3390/en15062218.

[24]. Djamali, A., Dossow, P., Hinterstocker, M., Schellinger, B., Sedlmeir, J., Völter, F., & Willburger, L. (2021). Asset logging in the energy sector: A scalable blockchain-based data platform. In Proceedings of the 10th DACH+ Conference on Energy Informatics (pp. 1-14). Virtual Conference.

[25]. Miao, S., Zhang, X., Asamoah, K.O., Gao, J., Qi, X. (2021). E-Chain: Blockchain-Based Energy Market for Smart Cities. In: Liu, Q., Liu, X., Shen, T., Qiu, X. (eds) The 10th International Conference on Computer Engineering and Networks. CENet 2020. Advances in Intelligent Systems and Computing, vol 1274. Springer, Singapore. https://doi.org/10.1007/978-981-15-8462-6_176.

[26]. R. B. Chakraborty, M. Pandey, and S. S. Rautaray, "Managing Computation Load on a Blockchain-based Multi-Layered Internet-of-Things Network", in Proc. Comput. Sci., vol. 132, pp. 469-476, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050918312400.

[27]. Y. Lin and C. Zhang, "A Method for Protecting Private Data in IPFS", in Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Nanjing, China, 2021, pp. 1-4. DOI: 10.1109/CSCWD49262.2021.9437830.

[28]. Fernando Rebollar Castelán. "Un Modelo Multicapas Basado en Blockchain que Fortalece la Integridad y Seguridad de Información Pública". Tesis doctoral, Doctorado en Ciencias de la Ingeniería, Universidad Autónoma del Estado de México, México, 2021.

[29]. CFE Esquema tarifario vigente. [Online]. Disponible en: https://app.cfe.mx/Aplicaciones/CCFE/Tarifas/TarifasCRENegocio/Negocio.aspx [Access: 4 may 2024].

[30]. Secretaría de Energía. "Avanza el proceso de separación de la CFE y su grupo corporativo". [Online]. Available: https://www.gob.mx/sener/prensa/avanza-el-proceso-de-separacion-de-la-cfe-y-su-grupo-corporativo. [Accessed: July 03, 2016].

[31]. Rousseau, "La reforma energética (2013-2014) a la luz de la nueva legislación sobre los impactos sociales de los proyectos", Foro int, vol. 60, no. 2, pp. 853-887, 2020. DOI: 10.24201/fi.v60i2.2740. [Online]. Available: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-013X2020000200853&lng=es&nrm=iso. Accessed on May 05, 2024.

[32]. Instituto Nacional de Electricidad y Energías Limpias, 'Reforma 113, Palmira, 62490 Morelos, Teléfono: (777) 3623811'. [En línea]. Disponible en: https://www.gob.mx/ineel/en#10052. [Access: 4 may 2024]."

Информация об авторах / Information about authors

Рене ГАРСИЯ-РЕЙЕС – аспирант последнего года на факультете программирования Мексиканского Национального центра исследований и технологических разработок (CENIDET). Его научные интересы включают децентрализованные технологии, криптографию, передовую измерительную инфраструктуру (AMI), науки о данных, методы добычи данных, Интернет энергии (IoE), облачные вычисления и визуализацию.

René GARCÍA-REYES – Final-year PhD student at the Computer Science Department of the National Center for Research and Technological Development (CENIDET) in Mexico. His research interests include decentralized technologies, cryptography, advanced metering infrastructure (AMI), data science, data mining, Internet of Energy (IoE), cloud computing, and virtualization.

Хавьер ОРТИС-ЭРНАНДЕС имеет степень PhD по автоматизации и программированию, штатный профессор факультета программирования Национального центра исследований и технологических разработок (Мексика). Сфера научных интересов: цифровая трансформация, инженерия требований, моделирование бизнес-процессов.

Javier ORTIZ-HERNANDEZ – PhD in Automation and Computer Science, full professor at the Computer Science Department of the National Center for Research and Technological Development in Mexico. Research interests: digital transformation, requirements engineering, business process modeling.

Рито МИХАРЕС имеет степень PhD по электрической и электронной инженерии. Профессор Мексиканского Национального института электроэнергии и чистой энергетики. Сфера научных интересов: цифровая обработка сигналов, встроенные системы, неразрушающее тестирование, ультраакустика.

Rito MIJAREZ – PhD in Electrical and Electronics Engineering, a full-time researcher at the Instituto Nacional de Electricidad y Energías Limpias in Mexico. Research interests: digital signal processing, embedded systems, non-destructive testing, ultrasonics.

Ясмин ЭРНАНДЕС имеет степень PhD по программированию от Технологического университета Монтеррей, профессор и факультета программирования Национального центра исследований и технологических разработок (Мексика). Научные интересы: искусственный интеллект, интеллектуальные обучающие системы, интеллектуальный анализ данных в образовании, аффективные вычисления, обработка естественного языка, человеко-машинное взаимодействие, машинное обучение.

Yasmín HERNÁNDEZ – PhD in Computer Science from Tecnológico de Monterrey and a full-time professor and researcher at the Computer Science Department of the National Center for Research and Technological Development (CENIDET) in Mexico. Her areas of expertise include artificial intelligence, intelligent tutoring systems, educational data mining, affective computing, natural language processing, human-computer interaction, and machine learning.

Хосе Альберто ЭРНАНДЕС-АГИЛАР имеет степень PhD по техническим и прикладным наукам от Центра исследований в области техники и прикладных наук при Автономном университете штата Морелос (UAEM). Член Мексиканской Национальной системы исследователей (SNI) 1-го уровня и является автором нескольких исследовательских и научно-популярных статей, а также трех книг. Его научные интересы включают интеллектуальный анализ данных, машинное обучение, глубокое обучение, алгоритмы оптимизации на графических процессорах (GPU) и прикладной искусственный интеллект.

José Alberto HERNÁNDEZ-AGUILAR – PhD of Engineering and Applied Sciences from the Center for Research in Engineering and Applied Sciences at the Autonomous University of the State of Morelos (UAEM). His research interests include data mining, machine learning, deep learning, optimization algorithms on Graphics Processing Units (GPUs), and applied artificial intelligence. Dr. Hernández Aguilar is a Level 1 member of the National System of Researchers (SNI) and has authored several research and outreach articles, as well as three books.

DOI: 10.15514/ISPRAS-2024-36(6)-11

Разработка и применение сервис-ориентированных научных приложений в инструментальном комплексе FDE-SWFs

[1] *А.Г. Феоктистов, ORCID: 0000-0002-9127-6162 <agf65@icc.ru>*
[1] *М.Л. Воскобойников, ORCID: 0000-0003-3034-4907 <mikev1988@mail.ru>*
[2,3] *А.Н. Черных, ORCID: 0000-0001-5029-5212 <chernykh@cicese.mx>*

[1] *Институт динамики систем и теории управления им. В.М. Матросова СО РАН, 664033, Иркутск, ул. Лермонтова, 134, а/я 29, Россия.*
[2] *Центр научных исследований и высшего образования Энсенада, Нижняя Калифорния, Мексика, 22860, Энсенада, шоссе Энсенада-Тихуана, 3918, почтовый ящик 360.*
[3] *Институт системного программирования им. В.П. Иванникова РАН, 109004, Москва, ул. А. Солженицына, 25, Россия.*

Аннотация. Рассматриваются актуальные аспекты организации сервис-ориентированных вычислений в вычислительной среде с гетерогенными ресурсами. Обсуждается развитие технологий разработки и применения сервис-ориентированных научных приложений, в которых схемы решения задач формируются в виде рабочих процессов. Приводятся существующие стандарты описания рабочих процессов. Предлагается новый инструментальный комплекс для создания сервис-ориентированных научных приложений, развивающий и дополняющий возможности систем подобного назначения.

Ключевые слова: вычислительная среда; гетерогенные ресурсы; сервис-ориентированные научные приложения; рабочие процессы.

Для цитирования: Феоктистов А.Г., Воскобойников М.Л., Черных А.Н. Разработка и применение сервис-ориентированных научных приложений в инструментальном комплексе FDE-SWFs. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 195–214. DOI: 10.15514/ISPRAS–2024–36(6)–11

Благодарности: Исследование проведено в рамках проекта № FWEW-2021-0005 «Технологии разработки и анализа предметно-ориентированных интеллектуальных систем группового управления в недетерминированных распределенных средах» (рег. № 121032400051-9) при поддержке Министерства науки и высшего образования Российской Федерации.

Development and Use of Service-Oriented Scientific Applications in the FDE-SWFs Toolkit

[1] *A.G. Feoktistov ORCID: 0000-0002-9127-6162 <agf@icc.ru>*
[1] *M.L. Voskoboinikov ORCID: 0000-0003-3034-4907 <mikev1988@mail.ru>*
[2,3] *A.N. Tchernykh ORCID: 0000-0001-5029-5212 <chernykh@cicese.mx>*

[1] *Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences, 134, Lermontov st., Irkutsk, postbox 292, 664033, Russia.*
[2] *Centro de investigación Científica y de educación Superior de Ensenada, Baja California, 22860, México, Ensenada, Ensenada-Tijuana Highway, 3918, postbox 360.*
[3] *Ivannikov Institute for System Programming of the Russian Academy of Sciences, 25, Alexander Solzhenitsyn st., Moscow, 109004, Russia.*

Abstract. The automation of workflow-based computing for solving large resource-intensive problems has undoubtedly had an impact on increasing the productivity of scientific research. In recent years, workflows have become the basis for abstractions covering data processing and high-performance computing using distributed applications. Workflow management systems are powerful tools for the collaborative development and use of distributed scientific applications. Nowadays, as part of the development of such systems, particular attention is currently being paid to supporting service-oriented scientific applications. Within this field of research, there is a large spectrum of problems related to the support of modular scientific applications, the standardization of their components and interfaces, the use of heterogeneous information and computing resources, and organization interdisciplinary research. Unfortunately, the solution to the above listed problems has not been fully implemented in known workflow management systems that support the development and use of service-oriented scientific applications. In this context, the paper discusses relevant aspects of organizing service-oriented computing in a computing environment with heterogeneous resources. The development of technologies for the development and use of service-oriented scientific applications, in which problem-solving schemes are formed in the form of workflows, is discussed. Existing standards for describing workflows are represented. A new framework for creating service-oriented scientific applications is proposed. It extends and complements the capabilities of systems for such purposes.

Keywords: heterogeneous distributed computing environment; heterogeneous resources; service-oriented scientific applications; workflows.

For citation: Feoktistov A.G., Voskoboinikov M.L., Tchernykh A.N. Development and use of service-oriented scientific applications in the FDE-SWFs toolkit. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 195-214 (in Russian). DOI: 10.15514/ISPRAS-2024-6(6)-11.

Acknowledgements. The study was supported by the Ministry of Science and Higher Education of the Russian Federation, project no. № FWEW-2021-0005 «Technologies for the development and analysis of subject-oriented intelligent group control systems in non-deterministic distributed environments».

1. Введение

Автоматизация расчетов на основе рабочих процессов (РП) в рамках решения больших ресурсоемких задач оказывает несомненное влияние на повышение продуктивности научных исследований. За последние годы РП стали основой абстракций, охватывающих обработку данных и проведение высокопроизводительных вычислений с помощью распределенных приложений. При этом использование специализированной системы управления рабочими процессами (СУРП, англ., Workflow Management System – WMS) зачастую освобождают конечных пользователей от необходимости вникать в детали выполнения РП, а также управления ими в вычислительной среде с гетерогенными ресурсами (ВСГР).

СУРП, такие как UNICORE [1], HTCondor [2], Pegasus [3] и другие программные комплексы [4-5], являются мощными инструментами командной разработки и применения

распределенных научных приложений. Они предназначены для интеграции программного обеспечения, планирования схем решения с его помощью научных и прикладных задач, назначения ресурсов для выполнения вычислений, запуска вычислительных процессов и управления ими, обработки данных и реализации других системных операций в распределенной программно-аппаратной среде.

В рамках развития СУРП особое внимание в настоящее время уделяется поддержке сервис-ориентированных научных приложений (СОНП) [5]. Сервис-ориентированное программирование направлено на разработку программных систем, поддерживающих взаимодействие приложений и сервисов различных типов на основе обмена сообщениями с использованием опубликованных и обнаруживаемых интерфейсов [6]. Зачастую сервисы предоставляют хороший способ реализовать дискретные операции приложений для вычислений и обработки данных в рамках бизнес-процессов в различных предметных областях исследований. Поэтому развитие сервис-ориентированных вычислений во многом обусловлено возможностью эффективного решения целого ряда проблем, связанных с поддержкой модульных научных приложений, стандартизации их компонентов и интерфейсов, использования разнородных информационных и вычислительных ресурсов, а также организации междисциплинарных исследований.

К сожалению, решение вышеперечисленных проблем в известных СУРП, поддерживающих СОНП (см., например, [7-9]), осуществлено не в полной мере. В этой связи в статье рассмотрены важные аспекты разработка и применение СОНП, а также представлен новый инструментальный комплекс, компоненты которого развивают и дополняют функционал известных в этом направлении СУРП.

2. Развитие сервис-ориентированных приложений

Парадигма разработки и применения СОНП представляет собой логическую эволюцию от объектно-ориентированных систем к системам сервисов. Как и в объектно-ориентированных системах, некоторыми фундаментальными концепциями веб-служб являются инкапсуляция, передача сообщений и динамическое связывание. Однако парадигма, основанная на сервисах, выходит за рамки сигнатур методов. Информация о функциях сервиса, его местоположении, способах доступа и др. также могут быть представлены в интерфейсе сервиса. Разработку СОНП также можно рассматривать как эволюцию модульного подхода к программированию, так как веб-сервисы представляют собой легкие, слабосвязанные, независимые от платформы и языка компоненты.

В настоящее время в основе рассматриваемой парадигмы доминирует сервис-ориентированная архитектура (СОА, англ., Service-Oriented Architecture – SOA). СОА основывается на использовании множества независимых веб-сервисов, выполняющих предопределенные операции, связанных с выполнением системных или прикладных приложений. Под веб-сервисом (англ., Web service) понимается программная система со стандартизированными интерфейсами, идентифицируемая уникальным веб-адресом (URL-адресом) [10]. При этом веб-сервисы не обладают знаниями о выполняемых приложениях, а приложения не нуждаются в информации о способах их выполнения веб-сервисами. Веб-технологии на основе СОА активно поддерживаются крупными компаниями-разработчиками, что обеспечивает их широкое распространение и использование.

Применительно к вычислительной среде СОА обуславливает ряд следующих важных преимуществ в процессах ее организации и применении [11]:

- многократное использование компонентов среды для построения сложных распределенных программных комплексов;
- модульный подход к разработке программного обеспечения (ПО);
- поддержку сетевого доступа к компонентам среды их разработчикам и пользователям, а также их взаимодействия между собой;

- обеспечение открытости среды за счет использования стандартов протоколов передачи данных и представления сервисных операций над этими данными;
- кроссплатформенность, позволяющую смягчить зависимость вычислительного процесса от используемых программно-аппаратных платформ и языков программирования;
- возможность безболезненной интеграции ПО разных разработчиков.

ПО, разработанное на основе СОА, как правило, реализуются в виде набора веб-служб, взаимодействующих по протоколу Simple Object Access Protocol (SOAP). Веб-служба является единицей модульности в рамках СОА ПО. В то же время СОА может быть реализована с использованием широкого спектра дополнительных технологий, таких как REpresentational State Transfer (REST), Remote Procedure Call (RPC), Distributed Component Object Model (DCOM) и Common Object Request Broker Architecture (CORBA).

Основными форматами представления структурированных данных являются eXtensible Markup Language (XML) и JavaScript Object Notation (JSON), для которых поддерживается проверка корректности данных с помощью XML Schema и JSON Schema соответственно. Неструктурированные данные, как правило, представляются в виде текстовых файлов или файлов других форматов. Передача данных между веб-сервисами осуществляется посредством их включения в тело сообщения (в случае небольшого размера передаваемой информации) или путем передачи адреса источника данных (англ., Uniform Resource Locator – URL), откуда их можно извекать (в случае большого размера данных).

Существуют разные способы описания сетевых служб или веб-сервисов [12]. В их числе можно выделить Web Service Description Language (WSDL) для описания веб-сервисов на основе SOAP и Web Application Description Language (WADL) для описания веб-приложений на основе HyperText Transfer Protocol (HTTP), в том числе веб-сервисов в стиле REST. В обоих случаях в качестве базового языка описания используется XML.

WSDL предназначен для описания веб-сервисов, доступа к ним и передаваемых между ними сообщений. Описание веб-сервиса на WSDL включает следующие основные разделы:

- определение типов данных, указывающих вид отправляемых и получаемых сервисом XML-сообщений, проверка которых осуществляется с помощью средств XML Schema;
- описание элементов данных – списка сообщений, используемых сервисом;
- задание абстрактных операций (портов) – списка методов, которые могут быть выполнены применительно к сообщениям;
- связывание сервисов – определение способов доставки сообщений;
- адрес вызова сервиса.

Последняя официальная спецификация языка WSDL Version 2.0 позволяет описывать, как вызовы различных специализированных веб-сервисов на основе SOAP, например, WPS-сервисов, так и сервисов на основе других протоколов, например, REST-сервисов.

Разработка и применение СОНП характеризуется рядом преимуществ по сравнению с другими типами приложений. Наличие в СОНП набора сервисов позволяет разработчикам приложения создавать, отлаживать, тестировать, разворачивать и модифицировать свои сервисы независимо от других разработчиков, что упрощает распределенную разработку приложения.

Каждый сервис может быть разработан и развернут на разных ресурсах с различными характеристиками их производительности, объемов оперативной и дисковой памяти, пропускной способностью интерконнекта и т. п. в Grid-системах, на ресурсах суперкомпьютеров или облачных платформ. С использованием контейнеризации сервисы можно запустить на нескольких параллельно работающих узлах без необходимости развертывания на новом узле всего приложения в целом.

Важным преимуществом СОНП является их отказоустойчивость. Отказ одного сервиса, как правило не приводит к отказу всего приложения в целом. При этом отказавший сервис может быть легко перезапущен или его операции могут быть выполнены другим сервисом при наличии вычислительной избыточности в СОНП.

В табл. 1 приведены сведения о разработках в области сервис-ориентированных вычислений. Рассматриваются следующие важные аспекты поддержки сервис-ориентированных вычислений на уровнях приложения и/или вычислительной среды в целом: разные типы сервисов, способы их спецификации, архитектура среды выполнения, сервис-ориентированные модели представления различных сущностей вычислительной среды, средства разработки, системы управления, СОНП и др.

Табл. 1. Разработки в области сервис-ориентированных вычислений.
Table 1. Developments in the field of service-oriented computing.

Источник / Source	Аспекты поддержки / Support aspects	Уровень поддержки / Support level
[13-17]	Grid и облачные вычисления, SOAP-сервисы, Grid-сервисы, SaaS, Globus Toolkit.	Среда
[9]	Облачные вычисления, SOAP-сервисы, СУРП, WaaS Cloud Platform	Приложение
[8]	Облачные вычисления, SOAP-сервисы, СУРП, HyperFlow	Приложение
[18]	Методика спецификации SOAP-сервисов, методика спецификации РП, WSDL, BPEL	Приложение
[19]	Микросервисы, композиции сервисов, синтез программ	Приложение
[20]	REST	Среда
[1]	Grid и облачные вычисления, Grid-сервисы, SOAP, REST, СУРП, UNICORE	Среда
[7]	Облачные вычисления, SOAP, REST, СУРП, Galaxy	Среда
[21]	Кооперативные вычисления, управление данными, микросервисы, iRODS	Среда, приложение
[22]	Облачные вычисления, SOAP, REST, MathCloud, Everest	Среда, приложение
[23]	Grid и облачные вычисления, SOAP-сервисы, Grid-сервисы, CAEBeans, испытательные стенды	Среда
[24, 25]	Облачные вычисления, РП, веб-сервисы, интеллектуализация управления вычислениями, IaaS, SaaS, PaaS, iPSE	Среда
[26]	СОНП	Приложение
[27]	HPC, Amazon Web Services, Google Compute Engine, OpenStack, Cloud Stack, IaaS, PaaS, SaaS	Среда
[28]	СОНП, обработка данных	Среда, приложение
[29]	СОНП, обработка данных	Среда, приложение
[30]	СОНП, обработка данных	Среда, приложение
[31]	Исследования в энергетике, СОНП, обработка данных	Среда
[32]	Геоинформатика, REST-сервисы, SOAP-сервисы, WPS-сервисы, композиции сервисов	Среда
[33]	REST-сервисы, SOAP-сервисы, микросервисы, композиции сервисов	Приложение
[34]	SOAP-сервисы, композиции сервисов, WSDL	Среда, приложение
[35]	Мультиагентное управление ресурсами, микросервисы, шаблоны сервисов	Среда

3. Стандарты представления РП

РП обеспечивают возможность проведения крупномасштабных научных экспериментов с использованием больших наборов данных. При этом процессы обработки данных распределяются на различные вычислительные ресурсы. РП могут включать операции

обнаружения и связывания ресурсов, а также сбора, обработки, анализа и визуализации данных. РП должны быть логичными, структурированными и надежными.

Операции РП выполняются в соответствии со схемой решения задачи в логической последовательности, определяемой его структурой. Применение стандартов для описания и выполнения РП позволяет распространять их среди научного сообщества и облегчает их многократное использование. РП могут быть размещены в общедоступных репозиториях.

Стандарты представления РП берут свое начало в области моделирования бизнес-процессов. Соответствующие решения были разработаны рядом коммерческих организаций, таких как IBM и Microsoft. Открытые стандарты разрабатываются независимыми консорциумами, в том числе The World Wide Web Consortium (W3C), Organization for the Advancement of Structured Information Standards (OASIS), Workflow Management Coalition (WFMC), Business Process Management Initiative (BPMI), United Nations Centre for Trade Facilitation and Electronic Business (UN/CEFACT) и Object Management Group (OMG) [36]. Некоторые консорциумы концентрируют усилия на разработке комплектов дополнительных стандартов, тогда как другие разрабатывают отдельные многоцелевые стандарты. Пока еще нет единого мнения относительно того, какие стандарты наиболее подходят для СОНП, а также не существует установленной структуры стандартов для СОА.

СУРП часто характеризуются описанием процессов с точки зрения потока данных, а не ориентации потока управления бизнес-процессом. В рамках ряда проектов проводятся исследования, сравнивающие применимость различных стандартов описания РП. Важным направлением является разработка стандартов, позволяющих учитывать требования к вычислениям и передаче данных для больших наборов данных, а также обеспечивающих разделение абстрактного уровня представления РП и уровня его выполнения на конкретных программно-аппаратных ресурсах. В целом успешная реализация РП зависит от использования системы стандартов, каждый из которых обеспечивает эффективное планирование и выполнение вычислительных операций и операций обработки данных.

Различными коммерческими организациями и консорциумами предложен ряд стандартов описания РП [37]. В том числе разработаны языки XML Process Definition Language (XPDL), XLANG, Web Services Flow Language (WSFL), Business Process Modeling Language (BPML), Business Process Specification Schema (BPSS), Web Services Conversation Languange (WSCL), Web Services Choreography Interface (WSCI), Yet Another Workflow Language (YAWL), Business Process Execution Language for Web Services (BPEL4WS или BPEL) 1.0, BPEL4WS 1.1, Web Services Choreography Description Language (WS-CDL) и Web Services Business Process Execution Language (WS-BPEL или BPEL) 2.0.

XPDL, разработанный Workflow Management Coalition (WFMC), предназначен для обмена определениями процессов между различными информационными системами, как в графическом, так и в семантическом виде. XPDL неоднократно пересматривался. Последняя ревизия состоялась в 2012 г.

XLANG от Microsoft является расширением WSDL. Его основное назначение заключается в определении бизнес-процессов и организации обмена сообщениями между веб-сервисами.

WSFL, разработанный компанией IBM, представляет собой XML-язык, описывающий бизнес-процесс в виде композиции веб-сервисов, в которой описывается последовательность вызовов операций сервисов. Порядок выполнения операций определяется на основе потоков управления и данных между сервисами. Бизнес-процесс определяет операции по получению, обработке и отмену данными в заданной последовательности.

BPML, предложенный BPMI, представляет собой язык описания бизнес-процессов на основе XML, представленный BPMI. Он предоставляет средства выполнения последовательных и параллельных операций, поддерживает ветвления и циклы, обеспечивает стандартные функции вызова сервисов, отправки и получения сообщений, позволяет разработчику РП

планировать выполнение задач в соответствии с заданным расписанием. В BPML предусмотрено управление РП с длительным сроком из выполнения.

BPSS представляет собой стандартную структуру, описывающую процесс обмена информацией. BPSS основана на метамодели UN/CEFACT. Эта схема позволяет предприятиям определять бизнес-транзакции и организовывать их сотрудничество между партнерами для электронного обмена документами и сигналами в коммерческих целях. BPSS входит в состав инструментария Electronic Business using XML (ebXML) от OASIS и UN/CEFACT.

WSCL от Hewlett-Packard предназначен для определения диалогов бизнес-уровня в виде общедоступных процессов, поддерживаемых веб-сервисами. WSCL определяет последовательность обмена XML-документами. Определения диалогов WSCL также являются XML-документами и поэтому могут интерпретироваться веб-сервисами.

YAWL представляет собой расширение XML и предназначен для формализованного описания бизнес-процессов. YAWL разработан в Техническом университете г. Эйндховен. Там также разработана специализированная программная платформа, поддерживающая текстовый и графичекий режимы построения бизнес-процессов и средства их выполнения. Исходный код программного обеспечения распространяется под лицензией GNU Lesser General Public License (LGPL).

WS-CDL создан W3C на основе XML для спецификации однорангового взаимодействия веб-сервисов на основе хореографии – упорядоченного обмена сообщениями между внешними сущностями. Спецификации сервисов определяют связи между гетерогенными вычислительными средами, используемыми для разработки и размещения веб-приложений. В целом обеспечение хореографии веб-сервисов позволяет организовать функционально совместимое одноранговое взаимодействие между любыми сервисами, независимо от поддерживающей платформы или модели программирования, используемой при их реализации.

К 2003 г. назрела необходимость перехода от разнородных стандартов отдельных консорциумов к некоторому единому стандарту. Усилия IBM, Microsoft, BEA, OASIS и др. консорциумов началась разработка и развитие BPEL. С развитием веб-сервисов произошло слияние WSFL и XLANG, что привело к появлению нового поколения языка спецификаций BPEL4WS 1.1. Язык BPEL4WS 1.1 позволил расширить модель взаимодействия веб-сервисов и сделать ее применимой для отображения бизнес-транзакций.

С появлением BPEL список стандартов, широко востребованных на практике, сокращается. Формально считается, что BPEL и XPDL обеспечивают оркестрацию взаимодействий между внутренними и внешними сущностями процессов, а WS-CDL и ebXML – хореаграфию. Однако функциональные возможности BPEL и XPDL позволяют описывать и хореографию.

В целом BPEL 2.0 [38] определяет модель для описания поведения РП в терминах взаимодействий (совокупности сообщений) между процессами и их партнерами (внешними сервисами). Важными дополнительными преимуществами BPEL являются следующие:

- РП могут не только вызывать веб-сервисы, но и сами быть представленными в виде сервисов;
- широкий спектр элементов управления и работы с данными, включающий элементы определения сложные структуры данных и параллельных процессов их обработки, циклы, ветвления, подпроцессы, элементы реализации асинхронного взаимодействия веб-сервисов и др.;
- использование WSDL для описания интерфейсов веб-сервисов обеспечивает гибкую интеграцию с другими РП и веб-приложениями;
- детальное описание РП реализует оркестровку внутренних и внешних сущностей процесса, а спецификация процесса обмена сообщениями отражает хореографию внешних сущностей (вызываемых веб-сервисов).

В дополнение к перечисленным выше стандартам неформальная рабочая группа, объединяющая представителей различных организаций и частных лиц, которые заинтересованны в переносимости рабочих процессов, разрабатывает язык Common Workflow Language (CWL) [39]. Цель работы группы заключается в создании спецификаций, позволяющих представителям научного сообщества описывать мощные, простые в использовании, портативные и поддерживающие воспроизводимость рабочие процессы. При представлении ряда конструкций рабочих процессов (например, конвейеров) CWL использует возможности языков YAML Ain't Markup Language (YAML) и JSON. Для контейнеризации прикладного ПО в портативных средах выполнения применяется система Docker. Он предназначен для описания РП с интенсивным использованием данных областях исследований, таких как биоинформатика, медицина, химия, физика и астрономия. Версия 1.0 языка CWL выпущена 8 июля 2016 г.

Разработка и развитие стандартов описания веб-сервисов представлено в ретроспективе на рис. 1. Представленное на данном рисунке развитие языков спецификации веб-сервисов обобщает, уточняет и дополняет диаграммы из [18].

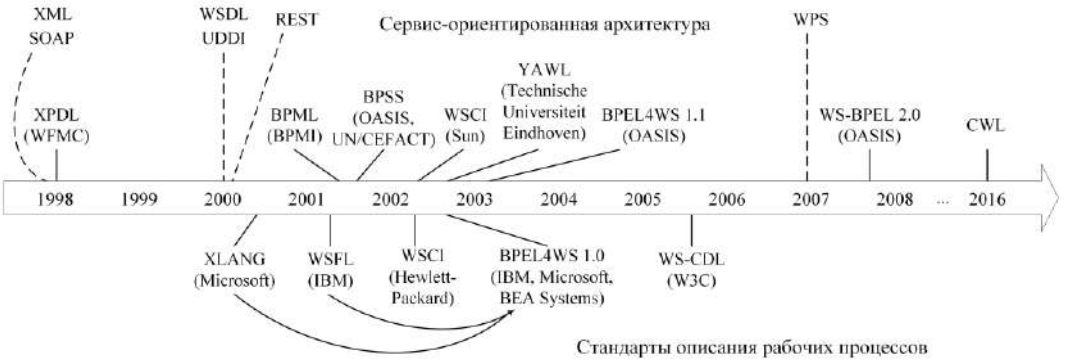


Рис. 1. Развитие языков спецификации рабочих процессов.
Fig. 1. Evolution of workflow specification languages.

4. Инструментальный комплекс

Существующие СУРП [40]) в том числе системы, обеспечивающие поддержку сервис-ориентированной парадигмы разработки прикладного ПО [7-9], широко используются для создания и применения научных приложений и управления ими в вычислительной среде. К сожалению, специализированные средства непрерывной интеграции разрабатываемого ПО, организации распределенных баз данных и работы с ними в памяти вычислительных узлов, тестирования и испытания компонентов СОНП в этих системах слабо развиты.

В этой связи в статье представлен новый инструментальный комплекс Framework for Development and Execution of Scientific WorkFlows (FDE-SWFs). FDE-SWFs относится к классу СУРП. Он базируется на подходе к разработке распределенных пакетов прикладных программ в ВСГР, поддерживаемом фреймворком Orlando Tools (OT) [41], и в тоже время существенно развивает и дополняет его функциональные возможности.

FDE-SWFs включает следующие основные подсистемы: пользовательский интерфейс, системные компоненты для конструирования и спецификации вычислительных моделей, а также управления вычислительными процессами и средой их выполнения. Для доступа пользователей к внешним информационно-вычислительным ресурсам и системам, с которыми им необходимо взаимодействовать при подготовке и проведении вычислительных экспериментов, он предоставляет набор специализированных API. База знаний инструментального комплекса содержит спецификацию вычислительных моделей

разрабатываемых приложений, РП и сведения о вычислительных ресурсах. Исходная информация и результаты выполнения РП хранятся в расчетных базах данных.

Вычислительная модель СОНП описывается в виде структуры

$$\mathrm{M} =< Z, T, M, O, S, P, SW, N, R >,$$

включающей следующие основные элементы:

- Z – множество значимых параметров предметной области приложения;
- T – множество их допустимых типов;
- M – множество программных модулей, представляющих алгоритмические знания предметной области;
- O – совокупность абстрактных вычислительных операций и операций обработки данных, отражающих семантику алгоритмических знаний в модели;
- S – множество сервисов, реализующих абстрактные операции;
- P – множество постановок задач, формулируемых в процедурной или непроцедурной формах на вычислительной модели;
- SW – множество РП, построенных на основе процедурных или непроцедурных постановок задач;
- N – множество ресурсов ВСГР, на которых размещаются и выполняются модули и сервисы приложения;
- R – связи между вышеперечисленными множествами объектов модели.

В общем случае множества Z, O, M, S и SW включают подмножества прикладных и системных объектов. Прикладные объекты создаются разработчиком приложения и дополняются предопределенными системными объектами, предназначенными для поддержки взаимодействия модулей, сервисов и РП с компонентами FDE-SWFs и внешними информационно-вычислительным ресурсами и системами при подготовке и проведении экспериментов. Конструктор вычислительной модели поддерживает возможность сознания новых операций в O на основе РП, а также генерацию программ на языке Python для РП с последующим включением сгенерированных программ в множество M.

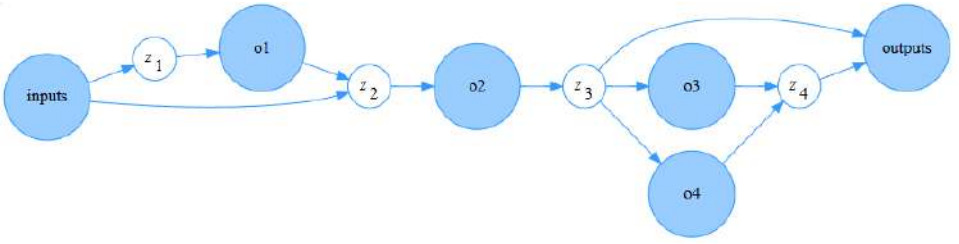
Вычислительная модель представляет концептуальный уровень вычислительной среды СОНП, на котором определяются понятия и связи между объектами программно-аппаратного, программно-алгоритмического и сервис-ориентированного уровней (рис. 2). Она позволяет разработчикам и пользователям приложений взаимодействовать с ВСГР и управлять ее компонентами на абстрактном концептуальном уровне, который во многом скрывает детали организации вычислительных процессов на других уровнях.

Конструктор РП обеспечивает их построение на основе процедурной или непроцедурной постановок задач. В первом случае, разработчик приложения самостоятельно строит рабочий процесс в интерактивном режиме с использованием предопределенного набора операторов выполнения операции, ветвления, различного вида циклов, обработки параллельных списков данных и других конструкций. В случае непроцедурной постановки задачи, разработчик определяет множества исходных и целевых параметров. Затем встроенный планировщик автоматически строит последовательность выполнения операций РП с целью расчета значений целевых параметров.

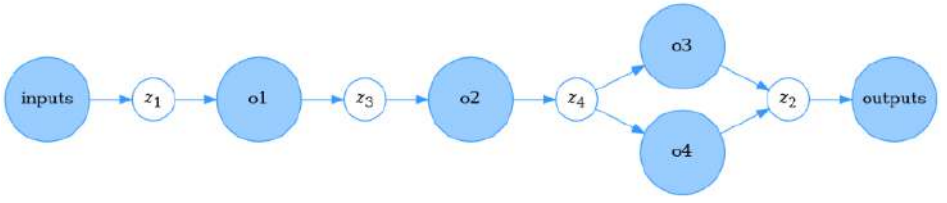
После завершения построения РП данный конструктор обеспечивает возможность автоматического формирования их спецификаций на языке BPEL, а также автоматическую генерацию программного кода выполнения этих рабочих процессов на языке Python с целью обеспечения возможности запуска и интерпретации РП автономно от среды FDE-SWFs.

Конструктор РП поддерживает визуализацию РП в виде двудольного направленного ациклического графа. Граф включает только два типа вершин (параметры и операции) и два типа дуг между этими вершинами: входная дуга соединяет вершину-входной параметр с вершиной-операцией, выходная дуга соединяет вершину-операцию с вершиной-выходным параметром. Примеры визуализации РП, построенных на основе процедурной и непроцедурной постановок задач, приведены на рис. 3.

Рис. 2. Уровни архитектуры вычислительной среды сервис-ориентированных приложений.
Fig. 2. Architecture layers of computing environment for service-oriented applications.



a)



b)

Рис. 3. Визуализация РП,
построенного на основе процедурной(a) и непроцедурной (b) постановок задач.
Fig. 3. Visualization of the scientific workflow
based on procedural (a) and non-procedural (b) problem formulations.

5. Система управления РП

На рис. 4 представлена схема работы ВСГР. FDE-SWFs реализован на программной платформе Node.js. Разработчики приложений с помощью веб-интерфейса комплекса описывают вычислительную модель СОНП. На вычислительной модели разработчик строит РП по процедурной или непроцедурной постановкам задач. РП, может быть реализован следующими способами: на основе композиции WSDL-сервисов, представленной на языке программирования Python; в виде отдельного WPS-сервиса, который реализуется традиционный РП на основе исполняемых модулей и заданий для внешних метапланировщиков и систем управления заданиями, таких как Condor DAGMan и HTCondor; с помощью стандартизированного декларативного языка BPEL.

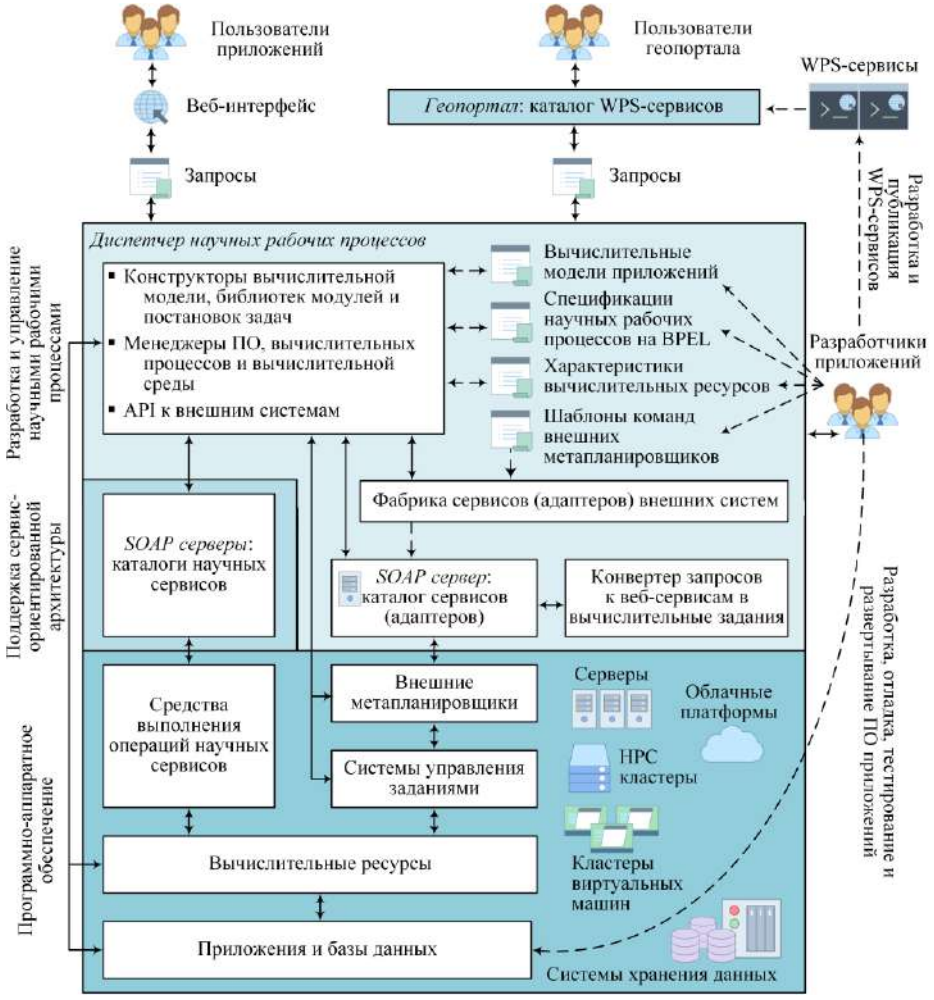


Рис. 4. Схема разработки и применения приложений.
Fig. 4. Scheme of application development and use.

В последнем случае, РП может использоваться любыми внешними СУРП, поддерживающими BPEL. На рис. 5 представлены схемы выполнения РП под управлением FDE-SWFs.

В рамках Схемы 1 пользователь с помощью интерфейса FDE-SWFs формулирует постановку задачи на вычислительную модель предметной области. По сформулированной постановке задачи модуль планирования вычислений выполняет построение плана решение задачи в виде РП во внутреннем представлении системы. Модуль преобразования производит конвертацию РП на BPEL для использования во внешних СУРП. Далее модуль интерпретации осуществляет асинхронно параллельное выполнение последовательности операций РП, представленных в виде композиции WSDL-сервисов. Полученные в результате выполнения композиции WSDL-сервисов расчетные данные сохраняются в хранилище данных. По завершению выполнения РП пользователь может осуществить визуализацию расчетных данных с целью их дальнейшего анализа.

Схема 2 в отличии от Схемы 1 предполагает, что операции РП, вместо представлены программными модулями. Процесс выполнения РП состоит в следующем. Специалист предметной области (конечный пользователь) формулирует постановку задачи, подготавливает исходную информацию и исполняемые модули с помощью менеджера вычислений. Затем выбирает необходимые ресурсы (ресурсы суперкомпьютерных ЦКП, облачных платформ, собственных высокопроизводительных серверов и др.) с учетом квот на их использование и способы доступа к ним. Менеджер автоматически строит РП и генерирует спецификацию вычислительного задания. Подсистема контейнеризации осуществляет выделение требуемых ресурсов, подготовку образов в соответствии с классами ресурсов и прикладными модулями РП с последующим запуском контейнеров на данных ресурсах (рис. 6). По готовности ресурсов для запуска РП передается сообщение СУПЗ (например, HTCondor или PBS Torque), которая также интегрирована в образы. После завершения выполнения РП метапланировщик отправляет полученные результаты в расчетную базу данных. Пользователь также может произвести визуализацию расчетных данных.

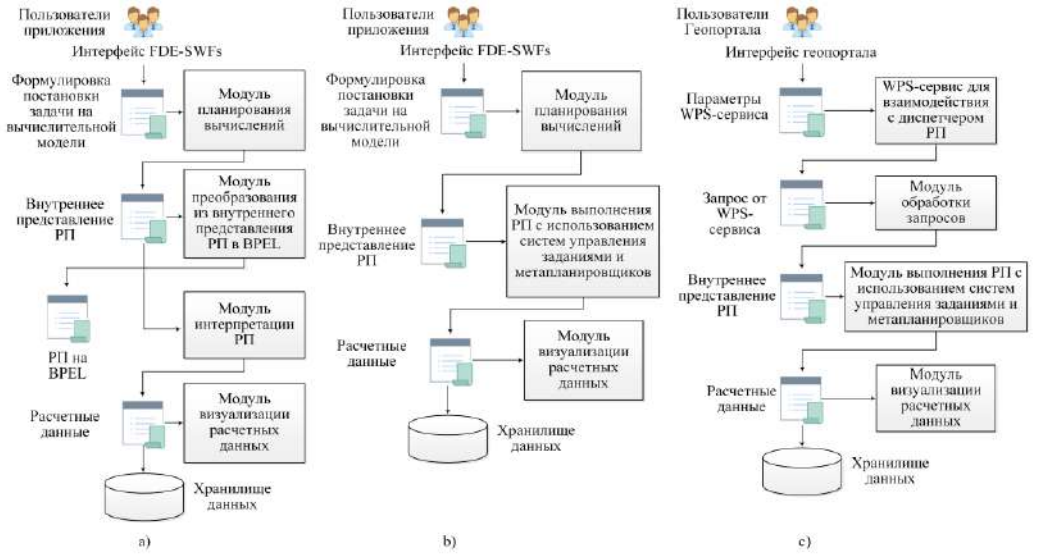


Рис. 5. Схемы выполнения РП: Схема 1 (a), Схема 2 (b) и Схема 3 (c).
Fig. 5. Schemes of the scientific workflow execution: Scheme 1 (a), Scheme 2 (b) and Scheme 3 (c).

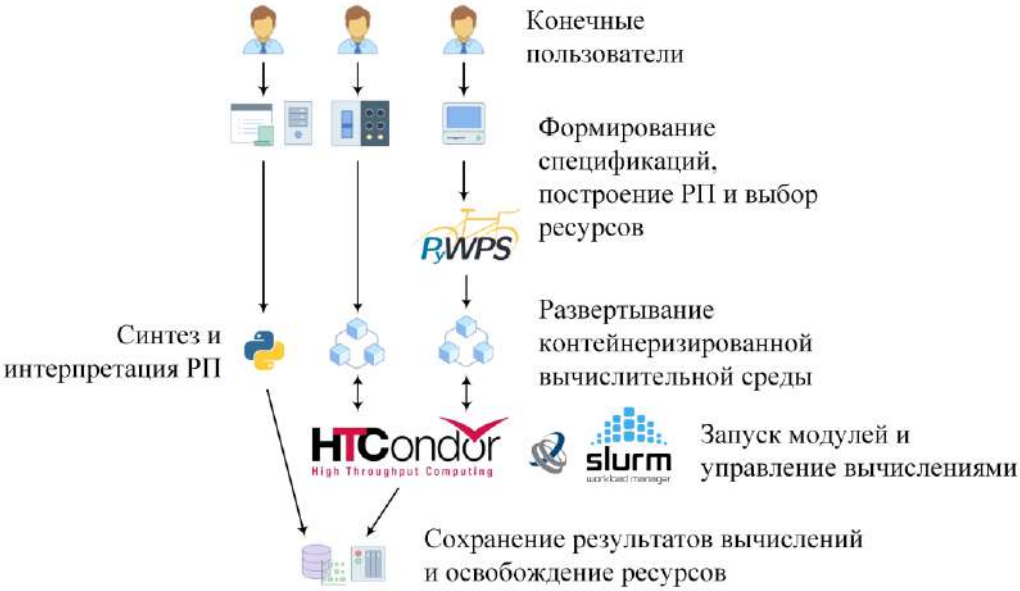


Рис. 6. Выполнение РП в контейнеризированной вычислительной среде.
Fig. 6. Scientific workflow execution in a containerized computing environment.

Возможность выполнения РП с помощью WPS-сервисов реализуется Схемой 3. В этом случае РП регистрируются как WPS-сервисы. FDE-SWFs автоматизирует создание, регистрацию и применение WPS-сервисов. В частности, FDE-SWFs автоматически регистрирует программные модули и РП в виде асинхронных WPS-сервисов в соответствующих каталогах на геопортале. РП могут включать вызовы других сервисов WPS, что позволяет работать с наборами сервисов. В FDE-SWFs поддерживается возможность обмена файлами между WPS-сервисами в качестве их параметров, в том числе обмен данными с системой хранения данных геопортала.

На рис. 7 приведен график времени разработки и выполнения тестового приложения с двумя РП в ОТ и FDE-SWFs для всех трех схем, представленных выше, с накоплением итогового результата. Были учтены следующие этапы разработки и применения приложения для одной из задач исследования процессов функционирования и свойств модели энергетической инфраструктуры:

1) описание вычислительной модели;
2) конструирование РП по процедурной постановке задачи;
3) конструирование РП по непроцедурной постановке задачи;
4) конфигурирование ресурсов ВСГР;
5) ввод исходных данных;
6) запуск и выполнение РП;
7) получение результатов вычислений;
8) визуализация результатов вычислений.

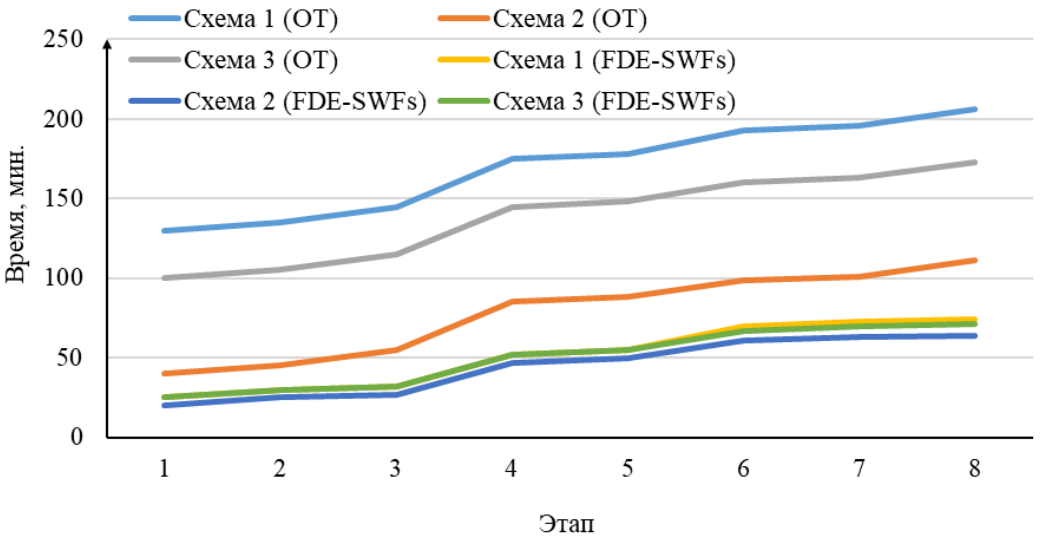


Рис. 7. Время разработки и выполнения тестового приложения.
Fig. 7. Test application development and execution makespan.

При работе с сервисами в ОТ дополнительно учитывались временные затраты на их создание в ручном режиме. В FDE-SWFs сервисы для выполнения модулей прикладного ПО создаются автоматически. Результаты сравнительного анализа времени выполнения показывают, что развитие функциональных возможностей ОТ в FDE-SWFs позволило существенно сократить временные затраты при выполнении всех трех схем.

6. Сравнительный анализ

В рамках сравнительного анализа рассматриваются следующие функциональные возможности FDE-SWFs:

- поддержка РП в виде направленного графа с циклами и ветвлениями (c_1);
- поддержка сервис-ориентированных РП (c_2);
- поддержка стандарта спецификации РП (c_3);
- поддержка взаимодействия с WPS-сервисами (c_4);
- поддержка контейнеризации ПО (c_5);
- генерация автономных программ на базовом языке программирования по выполнению РП независимо от среды СУРП (c_6).

Степень реализации функциональной возможности в СУРП определяется следующими показателями: реализовано (1), частично реализовано (0.5), не реализовано (0). Веса w_1–w_6 востребованности функциональной возможности заданы на основе агрегированных субъективных оценок пользователей. Результаты сравнительного анализа приведены в табл. 2. E – это суммарная оценка рассматриваемого набора функциональных возможностей СУРП, определяемая соотношением

$$E = \sum_{i=1}^{n} w_i c_i,$$

где n – число функциональных возможностей.

Табл. 2. Функциональные возможности СУРП.
Table 2. WMSs capabilities.

СУРП / WMS	c_1	c_2	c_3	c_4	c_5	c_6	w_1	w_2	w_3	w_4	w_5	w_6	E
OT	0.5	0	0	0.5	0.5	0	0.50	0.75	0.50	0.30	0.90	0.30	0.85
Pegasus	0	0	0	0	1.0	0.5						1.05	
Apache Airflow	0	1.0	0	0	1.0	0.5						1.80	
Galaxy	1.0	1.0	1.0	0	1.0	0						2.65	
FDE-SWFs	1.0	1.0	1.0	0.5	0.5	1.0						2.65	

Для сравнения выбраны предшественник FDE-SWFs – фреймворк OT, система Pegasus, одна из лидеров среди традиционных СУРП, и системы Apache Airflow [42] и Galaxy, представляющие активно развивающееся направление систем поддержки СОНП. В контексте рассматриваемого набора возможностей лидерами являются Galaxy и FDE-SWFs. Это во многом обусловлено поддержкой сервис-ориентированных РП, стандартизацией спецификаций рабочих процессов и особенностями специализации данных систем. Так, например, характеристики c_4 и c_6 имеют невысокие веса w_4 и w_6 среди пользователей традиционных СУРП. Однако с точки зрения пользователей FDE-SWFs в области исследования систем энергетики эти характеристики приобретают большую важность.

7. Практическое использование

В настоящее время FDE-SWFs успешно используется при разработке и применении ряда приложений для решения задач в области исследования свойств и процессов функционирования критически важных энергетических инфраструктур. В частности, с его помощью осуществляется подготовка и проведение крупномасштабных экспериментов для решения следующих задач:

- глобальный анализ степени уязвимости энергетических инфраструктур [43];
- оценка падения производительности систем энергетики из-за неординарного потока отказов их элементов при возникновении крупных внешних возмущений [43];
- профилирование и последующая оценка эффективности работы алгоритмов определения надежности функционирования энергетических систем различной конфигурации [44];
- определение наиболее подходящих алгоритмов структурно-параметрической оптимизации моделей энергетических комплексов на разных уровнях их территориально-отраслевой иерархии с помощью тестирования и многокритериального выбора исследуемых алгоритмов [45].

Отличительными особенностями FDE-SWFs при разработке СОНП на основе РП в сравнении с известными СУРП являются:

- использование технологии In-Memory Data Grid для размещения распределенных баз данных в оперативной памяти узлов среды с целью существенного ускорения обработки и анализа этих данных [43];
- создание испытательных стендов, предоставляющих разработчикам средства проведения экспериментов по оценке качества обрабатываемых данных, функционирования прикладного ПО, анализа результатов расчетов и других особенностей разрабатываемых приложений [44, 45];
- конструирование РП с использованием специальных системных операторов, в том числе операторов агрегирования и дезагрегирования данных, проведения многометодных расчетов, динамического планирования вычислений и др. [44, 45].

FDE-SWFs активно используется в учебном процессе в рамках дисциплин по изучению параллельных и распределенных вычислений для студентов и аспирантов образовательных организаций.

8. Заключение

С развитием научных приложений, основанных на рабочих процессах, возникает обоснованная потребность в их развертывании с использованием сервисов. Сервис-ориентированные СУРП зачастую существенно расширяют возможности традиционных систем подобного назначения. Они реализуют новую архитектуру, отвечающую современным бизнес-парадигмам проведения крупномасштабных междисциплинарных исследований на основе РП и в полной мере использующей преимущества информационных и коммуникационных технологий.

В этой связи в статье рассмотрены важные аспекты разработки и применения СОНП и реализации вышеупомянутой архитектуры в ВСГР средствами FDE-SWFs. Отличительная особенность FDE-SWFs заключается в обеспечении разнообразных способов построения и выполнения РП. РП строится по процедурной или непроцедурной постановкам задач. Они могут быть реализованы на основе композиции WSDL-сервисов, в виде отдельного WPS-сервиса, который реализуется традиционный РП на основе исполняемых модулей и заданий для внешних метапланировщиков и СУПЗ, а также с помощью стандартизированного декларативного языка BPEL. Соответственно в FDE-SWFs реализовано три схемы разработки и выполнения РП. Кроме того, показано, что в сравнении с известными сервис-ориентированными СУРП FDE-SWFs предоставляет ряд важных дополнительных возможностей. В их числе использование технологии In-Memory Data Grid, создание испытательных стендов и конструирование РП с использованием специальных системных операторов. В настоящее время FDE-SWFs успешно используется при разработке ряда приложений для решения задач анализа производительности и уязвимости энергетических инфраструктур и исследования эффективности работы алгоритмов структурно-параметрической оптимизации таких инфраструктур.

В рамках будущих исследований пристальное внимание будет вопросам повышения степени автоматизации интеграции и контейнеризации системного и прикладного ПО с целью дальнейшего улучшения эффективности развертывания и выполнения приложений в ВСГР, а также оптимизации использования вычислительных ресурсов. Кроме того, планируется разработка методов и средств обеспечения воспроизводимости вычислительных экспериментов по выполнению РП.

Список литературы / References

[1]. Erwin D.W., Snelling D.F. UNICORE: A Grid Computing Environment. Lecture Notes in Computer Science, 2001, vol. 2150, pp. 825-834. DOI: 10.1007/3-540-44681-8_116.

[2]. Litzkow M.J., Livny M., Mutka M.W. Condor – A hunter of idle workstations. In Proc. of the 8th International Conference on Distributed Computing Systems, Institute of Electrical and Electronics Engineers, 1988, pp. 104-111. DOI: 10.1109/DCS.1988.12507.

[3]. Deelman E., Vahi K., Juve G., Rynge M., Callaghan S., Maechling P. J., Mayani R., Chen W., da Silva R.F., Livny M., Wenger K. Pegasus, a workflow management system for science automation. Future Generation Computer Systems, 2015, vol. 46, pp. 17-35. DOI: 10.1016/j.future.2014.10.008.

[4]. Talia D. Workflow systems for science: Concepts and tools. International Scholarly Research Notices, 2013, vol. 2013, pp. 404525. DOI: 10.1155/2013/404525.

[5]. Da Silva R.F., Filgueira R., Pietri I., Jiang M., Sakellariou R., Deelman E. A characterization of workflow management systems for extreme-scale applications. Future Generation Computer Systems, 2017, vol. 75, pp. 228-238. DOI: 10.1016/j.future.2017.02.026.

[6]. Brown A., Johnston S., Kelly K. Using service-oriented architecture and component-based development to build web service applications. Rational Software Corporation, 2002, vol. 6. 16 p.

[7]. Afgan E., Baker D., Coraor N., Chapman B., Nekrutenko A., Taylor J. Galaxy CloudMan: delivering cloud compute clusters. BMC bioinformatics, 2010, vol. 11, no. 12, pp. 1-6. DOI: 10.1186/1471-2105-11-S12- S4.

[8]. Balis B. HyperFlow: A model of computation, programming approach and enactment engine for complex distributed workflows. Future Generation Computer Systems, 2016, vol. 55, pp. 147-162. DOI: 10.1016/j.future.2015.08.015.

[9]. Hilman M.H., Rodriguez M.A., Buyya R. Workflow-as-a-service cloud platform and deployment of bioinformatics workflow applications. In Knowledge Management in the Development of Data-Intensive Systems (In Mistrik I., Galster M., Maxim B., Tekiner-dogan B. editors). Boca Raton, FL, USA, CRC Press, 2021, pp. 205-226.

[10]. Papazoglou M. Web Services: Principles and Technology. New York, Pearson Education, 2008. 752 p.

[11]. Welke R., Hirschheim R., Schwarz A. Service-oriented architecture maturity. Computer, 2011, vol. 44, no. 2, pp. 61-67. DOI: 10.1109/MC.2011.56.

[12]. Tsalgatidou A., Pilioura T. An overview of standards and related technology in web services. Distributed and Parallel Databases, 2002, vol. 12, pp. 135-162. DOI: 10.1023/A:1016599017660.

[13]. Ananthakrishnan R., Chard K., Foster I., Tuecke S. Globus platform-as-a-service for collaborative science applications. Concurrency and Computation: Practice and Experience, 2015, vol. 27, no. 2, pp. 290-305. DOI: 10.1002/cpe.3262.

[14]. Foster I. Globus Online: Accelerating and democratizing science through cloud-based services. IEEE Internet Computing, 2011, vol. 15, no. 3, pp. 70-73. DOI: 10.1109/MIC.2011.64.

[15]. Foster I. Globus toolkit version 4: Software for service-oriented systems. Journal of Computer Science and Technology, 2006, vol. 21, pp. 513-520. DOI: 10.1007/s11390-006-0513-y.

[16]. Foster I., Kesselman C. The Grid: Blueprint for a new computing infrastructure. Morgan-Kaufmann, 2002. 667 p.

[17]. Foster I., Kesselman C. Globus: A metacomputing infrastructure toolkit. International Journal of Supercomputer Applications and High Performance Computing, 1997, vol. 11, no. 2, pp. 115-128. DOI: 10.1177/109434209701100205.

[18]. Juric M.B., Chandrasekaran S., Frece A., Hertis M., Srdic G. WS-BPEL 2.0 for SOA Composite applications with oracle SOA Suite 11g. Packt Publishing Ltd, 2010. 644 p.

[19]. Kim S., Bastani F.B., Yen I.L., Chen I.-R. High-assurance synthesis of security services from basic microservices. In Proc. of the 14th Intern. Symposium on Software Reliability Engineering (ISSRE 2003), IEEE, 2003, pp. 154-165. DOI: 10.1109/ISSRE.2003.1251039.

[20]. Fielding R.T. Architectural styles and the design of network-based software architectures. University of California, Irvine, 2000. 180 p.

[21]. Rajasekar A. iRODS primer: integrated rule-oriented data system. Morgan & Claypool Publishers, 2010. 143 p.

[22]. Sukhoroslov O. Building web-based services for practical exercises in parallel and distributed computing. Journal of Parallel and Distributed Computing, 2018, vol. 118, pp. 177-188. DOI: https://doi.org/10.1016/j.jpdc.2018.02.024.

[23]. Savchenko D.I., Radchenko G.I., Taipale O. Microservices validation: Mjolnirr platform case study. In Proc. of the 38th International convention on information and communication technology, electronics and microelectronics (MIPRO), IEEE, 2015, pp. 235-240. DOI: 10.1109/MIPRO.2015.7160271.

[24]. Бухановский А.В., Ковальчук С.В., Марьин С.В. Интеллектуальные программные комплексы компьютерного моделирования сложных систем: концепция, архитектура и примеры реализации. Известия высших учебных заведений. Приборостроение, 2009 г., том 52, № 10, стр. 5-24. (in Russian).

[25]. Васильев В.Н., Князьков К.В., Чуров Т.Н., Насонов Д.А., Марьин С.В., Ковальчук С.В., Бухановский А.В. CLAVIRE: облачная платформа для обработки данных больших объемов. Информационно-измерительные и управляющие системы, 2012, том 10, вып. 11, стр. 7-16. (in Russian).

[26]. Поляков С.В., Выродов А.В., Пузырьков Д.В., Якобовский М.В. Облачный сервис для решения многомасштабных задач нанотехнологии на суперкомпьютерных системах. Труды ИСП РАН, 2015, том 27, вып. 6, стр. 409-420. (in Russian). DOI 10.18522/2311-3103-2016-12-103114.

[27]. Кудрявцев А.О., Кошелев В.К., Избышев А.О., Дудина И.А., Курмангалеев Ш.Ф., Аветисян А.И., Иванников В.П., Велихов В.Е., Рябинкин Е.А. Разработка и реализация облачной системы для решения высокопроизводительных задач. Труды ИСП РАН, 2013, том 24, стр. 13-34.

[28]. Ханчук А.И., Сорокин А.А., Смагин С.И., Королев С.П., Макогонов С.В., Тарасов А.Г., Шестаков Н.В. Развитие информационно-телекоммуникационных систем в ДВО РАН. Информационные технологии и вычислительные системы, 2013, № 4, стр. 45-57.

[29]. Шокин Ю.И., Федотов А.М., Жижимов О.Л. Технологии создания распределенных информационных систем для поддержки научных исследований. Вычислительные технологии, 2015, том. 20, № 5, стр. 251-274.

[30]. Брагинская Л.П., Григорюк А.П., Ковалевский В.В., Загорулько Г.Б. Разработка научной среды для комплексных исследований в активной сейсмологии. Информационные технологии в науке, образовании и управлении: Материалы XLIV междунар. конф. и XIV междунар. конф. молодых учёных IT + S&E`16 (22 мая – 01 июня 2016 г., г. Гурзуф, Россия), М., Институт новых информационных технологий, 2016, стр. 3-10.

[31]. Массель Л.В., Болдырев Е.А., Макагонова Н.Н., Копайгородский А.Н., Черноусов А.В. ИТ-инфраструктура научных исследований: методический подход и реализация. Вычислительные технологии, 2006, том 11, № S8, стр. 59-68.

[32]. Бычков И.В., Ружников Г.М., Фёдоров Р.К., Шумилов А.С. Система планирования и выполнения композиций веб-сервисов в гетерогенной динамической среде. Вычислительные технологии, 2016, том 21, № 6, стр. 18-35.

[33]. Бычков И.В., Опарин Г.А., Феоктистов А.Г., Богданова В.Г., Пашинин А.А. Сервис-ориентированное управление распределенными вычислениями на основе мультиагентных технологий. Известия Южного федерального университета. Технические науки, 2014, № 12, стр. 17-27.

[34]. Бычков И.В., Опарин Г.А., Феоктистов А.Г., Богданова В.Г., Корсуков А.С. Сервис-ориентированный подход к организации распределённых вычислений с помощью инструментального комплекса DISCENT. Информационные технологии и вычислительные системы, 2014, № 2, стр. 7–15.

[35]. Kostromin R., Basharina O., Feoktistov A., Sidorov I. Microservice-Based Approach to Simulating Environmentally-Friendly Equipment of Infrastructure Objects Taking into Account Meteorological Data. Atmosphere, 2021, vol. 12, no. 9, pp. 1217. DOI: 10.3390/atmos12091217.

[36]. Михеев А., Орлов М. Война стандартов в мире workflow. Режим доступа: https://ecm-journal.ru/material/Vojjna-standartov-v-mire-workflow (дата обращения: 03.05.2024).

[37]. Артамонов И.В. Описание бизнес-процессов: вопросы стандартизации. Прикладная информатика, 2011, № 3 (33), стр. 20-28.

[38]. Web Services Business Process Execution Language Version 2.0. Available at: https://docs.oasis-open.org/wsbpel/2.0/wsbpel-v2.0.pdf, accessed 03.05.2024.

[39]. Common Workflow Language (CWL). Available at: https://www.commonwl.org, accessed 03.05.2024.

[40]. Feoktistov A., Edelev A., Tchernykh A., Gorsky S., Basharina O., Fereferov E. An Approach to Implementing High-Performance Computing for Problem Solving in Workflow-based Energy Infrastructure Resilience Studies. Computation, 2023, vol. 11, no. 12, pp. 243. DOI: 10.3390/computation11120243.

[41]. Feoktistov A., Gorsky S., Sidorov I., Bychkov I., Tchernykh A., Edelev A. Collaborative Development and Use of Scientific Applications in Orlando Tools: Integration, Delivery, and Deployment. Communications in Computer and Information Science, 2020, vol. 1087, pp. 18-32. DOI: 10.1007/978-3-030-41005-6_2.

[42]. Apache Airflow. Available at: https://airflow.apache.org/, accessed 03.05.2024.

[43]. Феоктистов А.Г., Костромин Р.О., Воскобойников М.Л., Ли-Дэ Д.И. Организация вычислительной среды разработки и применения научных рабочих процессов на основе контейнеризации. Вычислительные технологии, 2023, том 28, № 6, стр. 151-164. DOI: 10.25743/ICT.2023.28.6.013.

[44]. Danilov G., Voskoboinikov M. Testbed-based approach to testing a library for evaluating network reliability algorithms. In Proc. of the International Workshop on Critical Infrastructures in the Digital Worl (IWCI-2024), March 14-20, 2024, Bolshoe Goloustno, Russia, Irkutsk, ESI SB RAS, 2024, pp. 3-4.

[45]. Феоктистов А.Г., Воскобойников М.Л., Еделев А.В. Методы и средства разработки и применения испытательного стенда сервис-ориентированных приложений. Материалы VI Международного семинара по информационным, вычислительным и управляющим системам для распределенных сред (ICCS-DE 2024, 1-5 июля 2024 г., г. Иркутск, Россия), Иркутск, Изд-во ИДСТУ СО РАН, 2024, стр. 164-168.

Информация об авторах / Information about authors

Александр Геннадьевич ФЕОКТИСТОВ – доктор технических наук, доцент, главный научный сотрудник, заместитель директора ИДСТУ СО РАН по научной работе. Области исследования: параллельные и распределенные вычисления, научные приложения, мультиагентные технологии, имитационное моделирование и испытательные стенды.

Alexander Gennadevich FEOKTISTOV – Dr. Sci. (Tech.), Associate Professor, Chief Researcher, Deputy Director of ISDCT SB RAS for scientific work. Fields of research: parallel and distributed computing, scientific applications, multi-agent technologies, simulation modeling, and testbeds.

Михаил Леонтьевич ВОСКОБОЙНИКОВ – младший научный сотрудник ИДСТУ СО РАН. Сфера научных интересов: параллельные и распределенные вычисления, сервис-ориентированное программирование и испытательные стенды.

Mikhail Leontevich VOSKOBOINIKOV – Junior researcher of ISDCT SB RAS. Research interests: parallel and distributed computing, service-oriented programming, and testbeds.

Андрей Николаевич ЧЕРНЫХ – доктор физико-математических наук, главный научный сотрудник ИСП РАН, профессор Центра научных исследований и высшего образования в Энсенаде, Нижняя Калифорния, Мексика, директор лаборатории параллельных вычислений. Области исследований: теоретическая и экспериментальная многоцелевая оптимизация распределения ресурсов в облачной среде, безопасность, теория расписаний в условиях неопределенности, эвристики и метаэвристики, адаптивное распределение ресурсов, а также машинное обучение с сохранением конфиденциальности и полностью гомоморфным шифрованием.

Andrei Nikolaevitch TCHERNYKH – Dr. Sci (Phys.-Math.), Adjunct Chief Researcher at the ISP RAS, Full Professor in Computer Science at CICESE Research Center, Ensenada, Baja California, Mexico, Director of Parallel Computing Laboratory. Fields of research: multiobjective resource optimization in a cloud environment, both theoretical and experimental, security, scheduling with uncertainty, heuristics and meta-heuristics, adaptive resource allocation, and privacy-preserving machine learning with fully homomorphic encryption.

DOI: 10.15514/ISPRAS-2024-36(6)-12

Learning Analytics in Higher Education: a Decade in Systematic Literature Review

[1] *A. Salas-Martínez, ORCID: 0000-0002-3839-0978 <zS22000343@estudiantes.uv.mx>*
[1] *A. Ramírez-Martinell, ORCID: 0000-0003-2370-4994 <albramirez@uv.mx>*
[2] *S. Martínez-Ramos, ORCID: 0009-0000-1048-4975 <doc-095@itsperote.edu.mx>*

[1] *Facultad de Estadística e Informática, Universidad Veracruzana, Xalapa, México.*
[2] *Ingeniería Informática, Instituto Tecnológico Nacional de México, Perote, México.*

Abstract. In the last decade, Learning Analytics (LA) has evolved in a positive way, considering that the term emerged in 2011 through the Society for Learning Analytics Research (SoLAR). This area of data analytics can be identified as a specialization of Educational Data Mining (EDM). LA emphasizes student learning outcomes. In addition to, a better understanding of student learning behavior and processes. While EDM focuses on helping teachers and students with the analysis of the learning process using popular data mining methods. The purpose of this research is to explore the first decade of work with the application of Learning Analytics in Higher Education Institutions (HEI) in the context of Tutoring Information Systems (TIS), with the intention of supporting institutions, teachers and students to decrease dropout rates. This article presents a systematic literature review (SLR) with 17 primary studies, comprised between 2014 and 2024. The findings reflect the use of LA in improving or optimizing learning using student academic history obtained through Learning Management Systems (LMS), noting the scarcity of works with a focus on tutoring or academic advising. Ultimately, a gap is opened to apply LA in HEI, with information from Institutional Tutoring Program (PIT), integrated with information from an LMS, to contribute to student permanence.

Keywords: learning analytics; tutoring; counseling; systematic literature review; higher education institutions.

For citation: Salas-Martínez A., Ramirez-Martinell A., Martínez-Ramos S. Learning analytics in higher education: a decade in systematic literature review. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 215-230. DOI: 10.15514/ISPRAS-2024-36(6)-12.

Acknowledgements. This research was funded by the Universidad Veracruzana through the Faculty of Statistics and Informatics, the coordination of the PhD in Computer Science and the National Council for the Humanities, Sciences and Technologies (CONACYT) through grant number 730418. We also thank the Instituto Tecnológico Superior de Perote for their support in the development of the research.

Учебная аналитика в высшей школе: десятилетие в систематическом обзоре литературы

[1] *А. Салас-Мартинес, ORCID: 0000-0002-3839-0978 <zS22000343@estudiantes.uv.mx>*
[1] *А. Рамирес-Мартинель, ORCID: 0000-0003-2370-4994 <albramirez@uv.mx>*
[2] *С. Мартинес-Рамос, ORCID: 0009-0000-1048-4975 <doc-095@itsperote.edu.mx>*
[1] *Университет Веракруса, факультет статистики и информатики, Халапа, Мексика.*
[2] *Национальный технологический институт Мексики, Компьютерная инженерия, Пероте, Мексика.*

Аннотация. Термин учебная аналитика (LA) был введен Обществом исследований в области аналитики обучения (SoLAR) в 2011 году, и в последнее десятилетие это направление развивалось самым позитивным образом. Эта область аналитики данных может быть определена как интеллектуальный анализ образовательных данных (EDM). В дополнение к тому, что LA позволяет лучше понять поведение и процессы обучения студентов, учебная аналитика также обеспечивает более четкое выявление результатов обучения студентов, в то время как EDM фокусируется на оказании помощи преподавателям и учащимся при анализе процесса обучения с использованием популярных методов интеллектуального анализа данных. Целью данного исследования является изучение первого десятилетия работы с применением методов учебной аналитики в высших учебных заведениях (HEI) в контексте обучающих информационных систем с целью поддержки учреждений, преподавателей и студентов для снижения показателей отсева. В этой статье представлен систематический обзор литературы с 17-ю первичными исследованиями, проведенными в период с 2014 по 2024 год. Полученные результаты отражают использование LA для улучшения или оптимизации обучения с использованием академической истории студентов, полученной с помощью систем управления обучением, и отмечают нехватку работ с акцентом на обучение или академическое консультирование. В конечном счете, вскрыта недостаточность применения методов LA в ВУЗах к информации от Программы институционального обучения, интегрированной с информацией от систем управления обучением, которые могли бы внести свой вклад в постоянство студентов.

Ключевые слова: учебная аналитика; преподавание; консультирование; систематический обзор литературы; высшие учебные заведения.

Для цитирования: Салас-Мартинес А., Рамирес-Мартинель А., Мартинес-Рамос С. Учебная аналитика в высшей школе: десятилетие в систематическом обзоре литературы. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 215–230 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–12.

Благодарности. Грантом № 730418 исследование финансировалось Университетом Веракруса через факультет статистики и информатики, аспирантуру в области компьютерных наук, а также Национальный совет по гуманитарным наукам, наукам и технологиям (CONACYT). Мы также благодарим Высший технологический институт Пероте за поддержку исследования.

1. Introduction

The use of *learning analytics (LA)* in the online learning environment has increased exponentially, because its application can help institutions, teachers and tutors with problems such as decision making and measurement of student success, considering the digital footprint that can be obtained from students in each *Higher Education Institution (HEI)*. Currently it is a reality to mention that Higher Education has been forced to retake the use of technological tools such as *Learning Management Systems (LMS)* due to the arrival of the COVID-19 pandemic in two thousand twenty, because although these tools are not new, there was a resistance to their use in an official way. However, LMS gained great importance after the pandemic, tools such as Google Classroom, Microsoft Teams and Moodle are some of the most implemented by HEI in the world [1].

When talking about LA, it is important to mention that it is an incipient term, which arises as a specialization of Educational Data Mining (EDM), which first appeared in 2005 at the first workshop

on Educational Data Mining [2], being in 2008 the 1st International Conference on Educational Data Mining, held in Montreal, Quebec. While LA arises in the summer of 2012, being in the *Society for Learning Analytics Research (SoLAR)* where they define it as: *"the measurement, collection, analysis and reporting of data about learners and their contexts, in order to understand and optimize learning and the environments in which it occurs"* [3]. Meanwhile, EDM is defined as *"An emerging discipline concerned with developing methods for exploring unique types of data that come from educational settings and using those methods to better understand the learner and the environments in which they learn"* [2]. Therefore, we can say that first, LA places greater emphasis on student learning outcomes, a better understanding of student learning behavior and processes, in addition to better educational research. Second EDM focuses on helping teachers and students with the analysis of the learning process using popular data mining methods.

Some of the areas that have been considered for applying these analytics are:

1. Create alerts for stakeholders: monitor students' academic progress to quickly and correctly identify negative student behaviors, such as lack of motivation, dropping out, etc;
2. Group/profile students: separate students into groups according to their individual characteristics, personality traits, preferred learning methods, and other considerations;
3. Predict student performance: by calculating a performance estimate of grades, knowledge or score.

However, it has been observed that there has been little exploration of the academic tutoring that HEI students receive.

It is worth emphasizing that tutoring is an institutional program that arises with the intention of supporting students in their academic, personal and professional processes during their education in any HEI in Mexico. Therefore, Institutions such as Normal Schools, Universities and Technological Institutions, have their respective tutoring programs dedicated to support students during their academic life. Thus, from these systems it is possible to identify diverse situations of each student, from economic, health, academic and social points of view, problems that can directly impact their academic performance and even cause failure and in the worst-case scenario, desertion [1]. Therefore, this article is an extended version of the research published by Salas et al. 2024 [1], where the updates of the last decade to perform SLR, the discussion achieved during the presentation of the same at the *11th International Conference on Research and Innovation in Software Engineering* (CONISOFT 2023), as well as the integration of new findings, are considered. The integration of new findings, are part of the content that can be found in this publication. It aims to provide the reader with the following points:

1. Tutoring Information Systems (TIS) that use LA;
2. Identify on whom these TIS are focused on;
3. What are the most common interests in TISs;
4. How can SITs be quantified and categorized;
5. How the IES use the LA for tutoring;
6. How do they interpret and visualize LA-based data in the IES;
7. What information has been analyzed in the SIT of the IES and what they used for the analysis.

2. Background and related work

For contextualization it is important to define that according to Siemens [4], LA is defined as *"The measurement, collection, analysis and reporting of data about learners and their contexts, in order to understand and optimize learning and the environments in which it occurs"*. With the implementation of LA it is possible to find out more hidden information about learners in online learning. For this reason, it plays a relevant role in online learning whose main interest is to identify problems with learning and

improve the learning environment.

The following are some of the papers that were reviewed prior to conducting SLR with the intention of getting into the context of the work on Learning Analytics at HEI. In a study conducted by a university in Korea, empirical validation of the effects of a *Learning Analytics Dashboard (LAD)* was sought the results of this study, it was observed that students who interacted with the LAD scored higher compared to those who did not use it [5]. These results mark a path with respect to the need to review the LAD with features that motivate and support students who have different levels of academic performance [5].

On the other hand, a second study proposes the question: How do we begin the institutional adoption of learning analytics? This question is a common one among faculty, administrators, and researchers who seek to conduct Learning Analytics (LA) [1]. In summary, this study builds on established models for the adoption of business analytics, showcases two projects conducted in Australia, to develop and evaluate approaches for LA adoption in HEI [3]. The focus of the study highlights the importance of the socio-technical nature of LA and the complexities relevant to adoption in HEI.

A third study aimed to investigate student expectations regarding the characteristics of learning analytics systems and the willingness to use these features in learning. It was an exploratory and qualitative study, applied to 20 university students, who were interviewed about their expectations about learning analytics features. The findings of the study were complemented with a quantitative study applied to 216 students [6]. As results it was found that students expect Learning Analytics functions to support their planning and organization of learning processes, as well as provide self-assessments, adaptive recommendations, and produce personalized analyses of their learning activities [7].

Among the studies conducted, there is also ¨*Learning Dashboard for Insights and Support during Study Advice (LISSA)*, a LAD designed, developed and evaluated in collaboration with advisors, which aims to facilitate communication between advisors and students through the visualization of qualifications that are available in the HEI. The study found that the dashboard supports the ongoing dialogue between the advisor and student, motivating students, activating the conversation and providing tools for personalization, depth and nuance to the advising session, providing information at the factual, imperative and reflective levels, and engaging those involved in an active role during the session [7].

In another study, the use of a *Learning Analytics Dashboard (LAD)* to inform the teaching of five university professors was investigated using qualitative inductive analysis to identify salient emergent themes. The results of the study showed that instructors did not always draw on analytics with specific questions, but rather with general areas of curiosity [1]. The findings were synthesized into an analytical model of instructor use that provides useful categories of activities for future study and support [3].

On the other hand, an empirical study was identified, whose objective was to analyze both intrinsic and extrinsic motivation factors perceived through LMS such as Moodle. Said study is based on the *self-determination theory (SDT)*, the findings of the study reveal that intrinsic and extrinsic motivation significantly influence the effectiveness of student-perceived learning and the improvement of academic performance [8].

Another of the studies reviewed, focuses on determining the motivation of students in LA context, in this study aims to perceive the state of student motivation at a high level of abstraction. The results showed that it is possible to perceive the state of student motivation at a high level of abstraction [9].

Finally, one study identified was student-oriented, providing information and promoting self-regulated learning. In this study, a LAD design aligned with SRL *(Self-Regulated Learning)* theory was created, which was called *My Learning Analytics (MyLA)* [10], which seeks to better understand how students use a learning analytics tool. The study consisted of performing a sequential analysis of student interactions with three different dashboard visualizations implemented in an LMS. The results of this study showed discriminatory patterns in the use of the dashboard between different levels of

academic performance and self-regulated learning, particularly for students with low performance and high levels of self-regulation. The finding of this study highlights the importance of differences in students' experience with a student-oriented dashboard and emphasizes that one type of dashboard does not fit all in the design of learning analytics tools [11].

3. Research method

The research process was initiated through a Systematic Literature Review (SLR) in order to use explicit and systematic procedures as opposed to traditional research. Therefore, it followed Kitchenham and Charters' [12] guidelines on SLR in software engineering and Zhang's guidelines [13] proposed under the concept of *'Quasi-Gold Standard (QGS)'* applied in the identification of relevant software engineering studies. It is worth mentioning that Kitchenham's methodology is based on three phases that are planning, conducting and documentation, to achieve the identification of relevant studies. However, in this research has been combined with the methodology proposed by Zhang, to provide greater rigor in the search for relevant studies, considering that Zhang contemplates the automatic search and manual search to identify relevant studies, in addition to applying a sensitivity and accuracy assessment that confirm the rigor of the SLR. In addition to considering some recommendations from the works [14-15] the review proposed in this paper is composed of two subsections: Planning and Conduction.

3.1 Planning

In this phase, the formulation of the research questions was carried out, the search process was established, and its description follows.

3.1.1 Research questions

To drive the review process, the following seven research questions were generated, where each question seeks to clarify the panorama on the application of Learning Analytics in Higher Education Institutions (HEI).

1. **[RQ1]** Are there Tutoring Information Systems (TIS) used by the HEI where LA is used?
2. **[RQ2]** In the approaches used, is was at the center the student at the center, teaching tasks or tutorial management?
3. **[RQ3]** What are the declared interests identified in the TIS in HEI?
4. **[RQ4]** How can the TIS used in HEI be quantified and categorized?
5. **[RQ5]** How does HEI use LA for TIS?
6. **[RQ6]** How does the TIS use in HEI interpret and visualize LA-based data?
7. **[RQ7]** What information has been analyzed from the TIS and what was used to analyze the information?

3.1.2 Search process

The search process in this article followed the *"Quasigold standard"* (QGS) strategy [13]. This process consists of five steps: 1) Identify related databases, 2) Establish the QGS, 3) Define or obtain the search string, 4) Perform the automatic search and 5) Evaluate the performance of the search. Each of the steps is described below in the context of the investigation.

1) **Identify related databases**
   In this phase, journals were selected for the manual search and databases (DB) digital libraries and indexing services for the automatic search to. The following six journals were considered: *Knowledge and Learning, Technology, Informatics in Human Behavior*, for the SLR decade update the following journals were added: *Information Development, International*

Journal of Instruction and Perspectives on Psychological Science, for their relevance to the topic to be addressed, as well as for the Educational Institutions participating in the edition, the impact factor they maintain and the periodicity of the journal. Six others, by automatic search, coverage, overlap and accessibility of libraries and search engines. The following were included: *IEEE Xplore, ACM Digital Library, Springer Link, ScienceDirect, Wiley Online Library and EBSCOhost Academic Search*; the selected databases are available in the information resources of the *Consorcio Nacional de Recursos de Información Cient´ifica y Tecnológica (CONRICyT)* provided by the Universidad.

2) **Establish the QGS**

In this phase, the inclusion and exclusion criteria are defined, after which a manual search is carried out in the previously selected journals, which consists of analyzing all the volumes and identifying the articles that meet the established criteria, in this update of the SLR the last decade is contemplated, from 2014 to February 2024 and the criteria of the first version of the SLR are maintained; Table 1 shows the criteria established in this Systematic Literature Review (SLR).

Table 1. Inclusion And Exclusion Criteria.

Inclusion		Exclusion	
Id	Description	Id	Description
IC1	Access to the publication is through National Consortium for Scientific and Technological Information Resources (CONRICyT) provided by the university.	EC1	The publication is an exact duplicate of a study obtained from another search engine
IC2	The publication date is from 2014 to February 2024	EC2	The publication is not in Spanish or English
IC3	The publication must be a research article on software, learning analytics and tutoring in higher education institutions (Journal Article)	EC3	The full text is restricted in the retrieved publications
IC4	The publication must reference at least two search terms	EC4	The publication is not applied in Higher Education Institutions
IC5	The publication must answer at least one research question		

3) **Define or obtain the search string**

At this point in the review, the terms *"Learning Analytics", "Tutoring System", "Academic advising", "Academic counseling"* and *"Higher Education Institutions"* were taken as reference. It should be mentioned that depending on the databases (DB) consulted, the search string was refined and adapted depending on the fields available in the advanced search of each database (DB) [1]. Table 2 shows the search string used in a general way in all the previously mentioned search engines.

Table 2. Search String Executed.

Search String
("Learning Analytics") AND ("Mentoring system" OR "Academic Advising" OR "Tutoring System" OR "Academic Counseling")

4) **Perform the automatic search**

In this phase, the search was performed in each of the database (DB) selected by applying the specific syntax in each of them. In this update, a total of 157 publications were obtained, achieving an increase of 86 studies linked to our search string, it notes, unlike the SLR presented previously [1], the 157 studies found were verified by applying the inclusion and

exclusion criteria to obtain the final corpus.

Selection of Primary Studies. To carry out the selection of these studies, it was necessary to apply the process of inclusion and exclusion criteria, which consisted of three stages, as shown in Table 3. This organization served to reduce the number of publications while retaining the relevant studies for subsequent analysis.

Table 3. Study Selection Process.

Stages	Criteria
Stage 1	IC1, IC2 and EC1
Stage 2	IC3, IC4, EC2 and EC3
Stage 3	IC5 and EC4

The following is the selection of candidate studies after application of the inclusion and exclusion criteria, as shown in Table 4.

Table 4. Results of the selection of candidate studies.

Id	Source	Candidate papers	Filter 1	Criteria I/E	Included
1	IEEE	21	0	17	4
2	ACM	22	4	15	3
3	Springer	28	0	16	12
4	ScienceDirect	82	67	7	9
5	Wiley OnLine	2	0	0	2
6	EBSCOhost	2	0	1	1
	Total	157	71	56	30

5) **Evaluate the performance of the search**

At this point the results of the automatic search are compared with the manual search *(QGS)*. To achieve this, we used the equations proposed by Zhang et al. [13]. First, the Equation sensitivity or recovery was calculated Eq. (1), to obtain the number of relevant studies retrieved, we subtract from the total number of studies retrieved automatically, which are 157. 140 studies were not relevant. To obtain the total number of relevant articles, we divided it by the total number of Relevant studies found, thus achieving 100% of the corpus ([157-140/17]*100). Afterwards, to calculate the precision we found 17 studies found by *QGS*. 10 are not relevant, therefore, we proceeded to use the Eq. (2), to obtain the precision, the number of relevant studies recovered, subtracting the 71 studies obtained through the automatic search, the 10 studies that are not relevant and then we divide it by the total number of studies retrieved in the automatic search, thus obtaining 93% of the corpus ([157-10/157]*100). Therefore, it is identified that both parameters are within the suggested threshold that indicates the percentages must be greater than 70% to be acceptable, the maximum Equation sensitivity can be observed in Eq. (3) and the optimal precision in Eq. (4).

$$Sensitivity = \frac{NRSR}{TNRS} 100\% \quad (1)$$

$$Precision = \frac{NRSR}{NRS} 100\% \quad (2)$$

where: *NRSR* = Number of relevant studies retrieved

TNRS = Total number of relevant studies

NSR = Number of studies retrieved

$$Sensitivity = \frac{157 - 140}{17} 100\% = 1.0 \quad (3)$$

$$Precision = \frac{157 - 10}{157} 100\% = 0.93 \quad (4)$$

3.2 Conduction

3.2.1 Quality assessment

At this point in of the process, the 30 selected studies are taken up again and again subjected to validation to identify only those studies that meet the necessary quality; therefore, the quality assessment instrument was prepared, which contains the quality control questions. A value of 1 was assigned to the questions that are answered with the word yes, a value of 0.5 for those questions that are considered to be partially compliant and 0 for those that are not. Table 5 shows the questions asked to assess the quality of the study.

All the studies found were evaluated with the proposed instrument to guarantee the quality of the chosen studies. The possible score to achieve was between 0 and 8 points. After the evaluation, all the studies that achieved a score greater or equal to 6.5 were considered; see Table 6. It was observed that 56% (17 studies). See Table 7 met the established quality criteria for the most part, while 43% (13 studies) did not meet them, therefore they were discarded from the final selection.

3.2.2 Data Extraction

This phase consisted of extracting the most relevant data from each of the primary studies identified, with the support of the *Parsifal* platform. Bibliographic information was extracted for each study such as: title, authors, year of publication, source, type of publication, DOI, keywords and abstract. It also includes information that helps to answer the research questions.

4. Results

This section of the article presents different results obtained during the SLR of the research, such as: text analysis applied to the bibliography of the corpus through the *VosViewer software*, frequent word cloud of the content of the articles generated from *MAXQDA 2022*, in addition to answering the research questions. A narrative synthesis is provided based on the data identified in the research corpus. To start, an overlay visualization generated by *VosViewer* is shown, see Fig. 1. It is possible to observe the grouping through four clusters formed with titles and summary of the corpus, where the terms *Learning Analytics, Educational data Mining, Environment and Dashboard* are found, demonstrating the linkage and relevance they present in the research. On the other hand, as shown in Fig. 2, relevant works within the SLR are scarce, however, the application of LA in HEI is a topic of interest from 2015 to date, observing a positive trend of articles towards 2024.

4.1 Answers to research questions

[RQ1] Are there TIS used by HEI where LA are used? Based in the work carried out by Chatti et al. [32] who proposes a reference model for Learning Analytics (LA) based on four specific dimensions, (1) ***what*** (e.g., data, environment and context), (2) ***why*** (e.g., objectives), (3) ***how*** (e.g., techniques/methods) and (4) ***who*** (e.g., stakeholders), model that helps to have an overview of LA and its concepts of relevance, in addition to including the review carried out by Bodily et al. [17] where they categorize the works under an approach between various subfields of educational technologies. In this review we analyze the objectives and technologies that guide those interested in making effective decisions about teaching, within the analysis we can identify the following categories to classify the most relevant jobs in the educational field and LA, see Table 8.

[RQ2] In the approaches used, is the student at the center, the teaching tasks or the management of tutorials? Based on the studies reviewed, it can be classified that the works where Learning Analytics (LA) is applied are mainly focused on the following actors [16]: teachers, students, tutors and researchers, with the latter having less presence in the research reviewed [25, 28]. According

to Robert Bodily [17] in his review of student-oriented learning analysis dashboards and educational recommender systems, most of the systems found are oriented 74 percent to the instructor, and he also states that researchers do not conduct much research on the impact of the systems on teaching and learning. Also, Perez Sánchez [27], Ranjeeth [23] and Rafique [26] focus on students.

Table 5. Quality Assessment Instrument Used for Included Study Evaluation.

Id	Question
QA01	Are the objectives, research questions, and hypotheses (if any) clear and relevant?
QA02	Is there an adequate description of the context in which the research was conducted?
QA03	Is the suitability of the case to address the research questions clearly motivated?
QA04	Are the case and its units of analysis well defined?
QA05	Is the case study based on theory or linked to existing literature?
QA06	Are the data collection procedures sufficient for the purpose of the case study (data sources, collection, validation)?
QA07	Are ethical issues (personal intentions, integrity, confidentiality, consent, review board approval) adequately addressed?
QA08	Is a clear chain of evidence established from observations to conclusions?

Table 6. Primary Studies Quality Assessment.

Id	QA1	QA2	QA3	QA4	QA5	QA6	QA7	QA8	Total
PS01	1	1	1	1	0.5	0.5	1	1	7.0
PS02	1	0.5	1	0.5	1	1	1	1	7.0
PS03	0.5	1	1	1	1	1	1	1	7.5
PS04	1	1	1	1	1	1	1	0.5	7.5
PS05	1	1	0.5	1	0.5	1	1	1	7
PS06	1	1	1	1	1	1	1	1	8.0
PS07	1	1	1	1	1	0.5	0.5	0.5	6.5
PS08	1	1	1	1	1	1	1	1	8
PS09	1	1	1	0.5	1	1	0.5	1	7.0
PS10	1	1	1	1	1	1	0.5	1	7.5
PS11	1	1	1	1	1	0.5	0.5	0.5	6.5
PS12	1	1	1	1	1	0.5	0.5	1	7.0
PS13	1	1	1	1	1	1	0.5	0.5	7.0
PS14	1	1	1	0.5	1	1	1	0.5	7.0
PS15	1	1	1	0.5	0.5	0.5	1	1	6.5
PS16	1	1	1	0.5	1	0.5	0.5	1	6.5
PS17	1	1	1	0.5	1	1	0.5	1	7.0

Table 7. Primary Studies.

Id	Author	Year	Database
PS01	Reyes et al. [16]	2015	Springer
PS02	Siemens et al. [4]	2015	ACM
PS03	Bodily et al. [17]	2017	IEEE
PS04	Rojas et al. [18]	2017	EBSCOhost
PS05	Tempelaar et al. [19]	2017	IEEE
PS06	Viberg et al. [20]	2018	ScienceDirect
PS07	Herodotou et al. [21]	2019	Springer
PS08	Guerra et al. [22]	2020	Wiley
PS09	Ranjeeth et al. [23]	2020	ScienceDirect
PS10	De Laet et al [24]	2020	Wiley
PS11	Guzman et al. [25]	2021	Springer
PS12	Rafique et al. [26]	2021	IEEE

PS13	Perez et al. [27]	2022	Springer
PS14	Kaliisa et al. [28]	2022	ScienceDirect
PS15	Hao et al. [29]	2023	ScienceDirect
PS16	Prinsloo et al. [30]	2023	Springer
PS17	Kaur et al. [31]	2023	IEEE

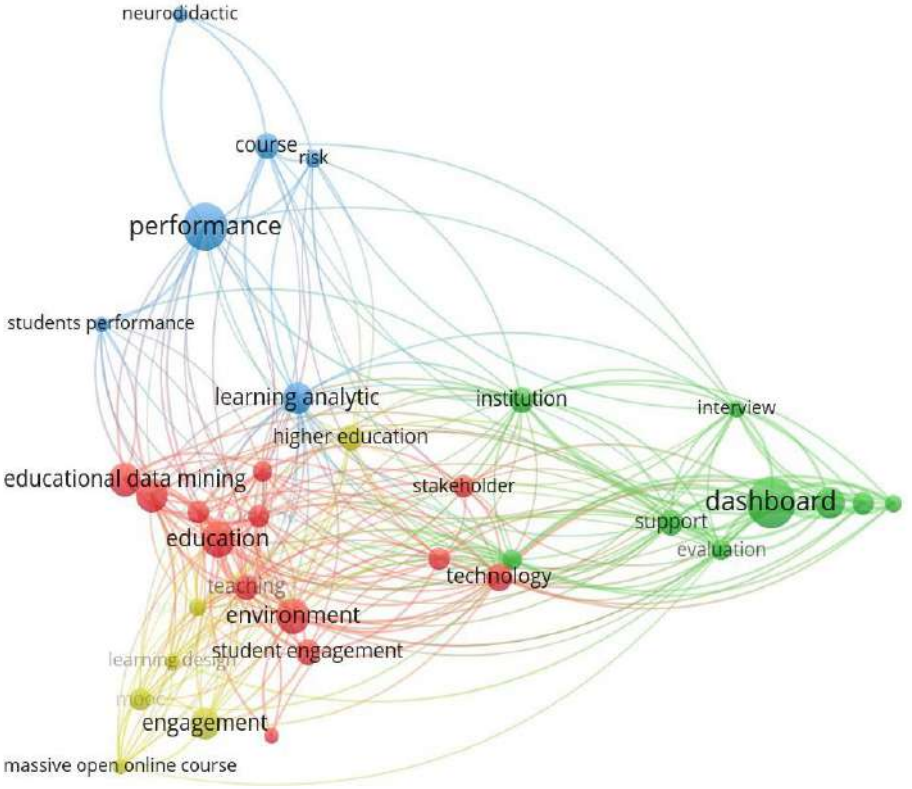


Fig. 1. An overlay visualization from research corpus.

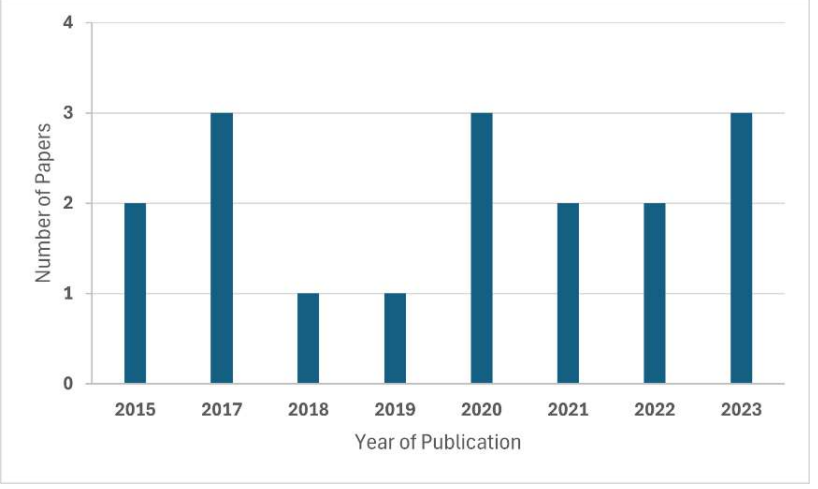


Fig. 2. Year of publication.

Table 8. Categories of systems using learning analytics in higher education institutions.

Category	Article
Intelligent tutorial systems	[17]
Predictive systems	[21], [26]
Academic performance system	[17],[20]
Educational recommendation systems	[27], [26]
Learning boards	[17], [22], [21], [26], [28], [20]
Educational data mining system	[17], [27]

For example, in Ranjeeth's literature study, some of the predictions that have been made are: predictions of student grade point average (GPA), prediction of student performance in graduate programs, prediction of instructor performance and likely student performance in gaining admission to college, prediction of attrition from college programs, and prediction of student grades using social network theory analysis. For his part Rafique says that student performance can be predicted from the student's digital fingerprints [22], i.e., demographics, behavior, facial emotion control records while using an intelligent tutoring system [21]. While for researchers the most predominant focus is on an online learning environment to predict student performance and timely intervention, however, Rafique expresses that it is very limited work in traditional learning environments.

[RQ3] What are the declared interests that are identified in the TIS in HEI? According to the research work conducted by Bodily [17] in 2017, it is possible to identify that there is an interest in identifying student-oriented LA reporting systems with respect to their purpose, functionality and the types of data collected. Also, Schwendimann et al. [17] mentions the interest in the mechanisms by which student-oriented systems attempt to improve teaching and learning, which requires analysis through different categories such as type of data, target users, and evaluation. Learning Analytics Dashboards (LAD) also identified by their acronym LAD were found to have evaluated categories such as: goal orientation, usefulness of information, visual effectiveness, ease of use, comprehension, reflection, motivation for learning, behavior change, performance improvement, and competency development.

[RQ4] How can the TIS used in HEI be quantified and categorized? In the various works reviewed, multiple approaches and objectives have been observed, however, none of them gives an answer to this question since most of the cases where tutoring is discussed, their focus is on intelligent tutoring systems, as identified by Bodily, to this point the only ones that come closest to working with tutoring in Latin America, are the systems generated by Learning Analytics in Latin America (LALA) expressed in the research of Guerra et al. [22]. This arises from a framework called COALA [24] (Context Adaptation for Learning Analytics), which is constituted by four dimensions for adapting tools: objectives of using a Learning Analytics Dashboards (LAD) (e.g., identifying subjects in which students have low or high performance), stakeholders (e.g., advisors, teachers, students and administrative staff), key moments in which the use occurs (e.g., at the beginning of the academic year, when a course is registered or when they receive grades) and the interaction of stakeholders (e.g., face-to-face sessions with the advisor-student). This project was conducted within the context of Latin America partner institutions, the participating institutions were *University of Cuenca in Ecuador* (Cuenca), *University Austral of Chile* (UACh) and *Polytechnic Superior School of the Litoral in Ecuador* (ESPOL) [24].

[RQ5] How do HEI use LA for TIS? Among the works reviewed, there are three cases where the use of LA applied to tutorial information systems has been most closely approached. Identified in the following institutions: *University of Cuenca, University Austral of Chile and Polytechnic Superior School of the Litoral.* See Table 9. The three cases coincide in combining information from the curricular structure and academic records to observe student progress [1]. However, the three Latin American universities adapted an advisory board, originally implemented at KU Leuven in Belgium. In all three cases, the

context was the main factor for adapting the dashboard, taking up that the LALA project [22] focuses on four different elements of the context such as: Objective, Actors, Key Moments and Interactions.

[RQ6] How do TIS used in HEI interpret and visualize the data based on LA? Several institutions have begun to adopt Predictive Learning Analytics (PLA) [21], they use a number of computational techniques (e.g., Bayesian modeling, cluster analysis, predictive modeling) to identify which students will pass a course and which are at risk. According to Merceron's categorization [20], it is identified that predictive methods (regression and classification) with 32% are considered the most frequent, below are relationship mining methods (association rules, correlations, sequential patterns and causal data mining) and methods for distilling data for human judgment tied with 24% frequency, where statistics and visualization are included. According to Viberg et al. [20] the application of methods for data analysis has been increasing from 2014 to 2017, reflecting from 2017 an increase in mining methods compared to previous years. In the work of Kim et al. [34], k-medoids clustering and Random Forest classification followed by logistic regression were applied for the analysis of the identified cluster profiles to analyze students' self-learning patterns in asynchronous mode. Likewise, to predict whether students pass or fail the course, the following models were used: Ramdom forest (RF), K-Nearest Neighbors (KNN), logistic regression (LR), neural networks (NNETS), TreeBagging (TB) and Bayesian additive regression trees (BART) Perez et al. [27].

Table 9. Some Applications of Learning Analytics in Higher Education Institutions.

Classification	Name	Goal	Author	Country
Learning Analytics Dashboard (LAD)	OUA dashboard (Open University Analyze)	Helps teachers identify at-risk students in online course	Herodotou et al. [21]	United Kingdom
	ESPOL LAD (Escuela Superior Politecnica del Litoral)	Support student-advisor dialogue when advising study plan in student	Guerra et al. [22]	Belgium
	LISSA (Learning dashboard for Insights and Support during Study Advice)	Support student–advisor dialogue focus on first-year students	Charleer et al. [7]	Belgium
	LADA (Learning Analytics Dashboard for Advisors)	Support advice on study plan by advisors	De Laet et al. [24]	Belgium
	AvAc (Advising dashboard Avance Academico)		Gutierrez et al. [33]	Cuenca
	TrAC dashboard (Trayectoria Academica y Curricular	Inspired by dashboard LISSA superimposes the academic records on the curricular structure, since this is the "natural" way in which academic progress is understood in the institution	Gutierrez et al. [33]	Ecuador
Framework	COALA framework (Context Adaptation for Learning Analytics)	To evaluate the support provided by the adapted dashboard comprises three modules including a visualization module, a module for group formation and intervention, and a prediction module	Guerra et al. [22]	Chile
	Smart Learning		Rafique et al. [26]	Pakistan
	SHEILA (Supporting Higher Education to Integrate LA)	The proposed framework will enhance systematic adoption of learning analytics on a wide scale	Viberg et al. [20]	United Kingdom

[RQ7] What information has been analyzed from the TIS and what did they use for the analysis of the information? It has been identified that to date only the data available in some of the educational platforms have been used to generate conclusions from the LA perspective. Among the data identified we can find: student demographics: age, gender, disability, previous grades, ethnicity, successful completion of previous courses, previous experience of the student at the university (new versus continuing student), best score in the previous course and sum of credits earned [21]. In the work of Perez et al. [27], characteristics such as: LMS, numbers of accesses, participation scores, learning activity ratings, submissions, published content, completed learning activities and peer reviews were contemplated, where these characteristics presented significant statistics between students who failed and those who passed.

5. Discussion

With the systematic review of literature, we have realized that there is little research that considers the tutoring received by students in higher education institutions, through student support programs, such as the Institutional Tutoring Program (PIT), which is promoted by institutions such as the UN (United Nations) and ANUIES (National Association of Universities and Higher Education Institutions) in Mexico. At the same time, it has been observed that when researchers make use of data, these are only data extracted from LMS, or application of surveys. Therefore, an important gap opens up for applying LA by integrating data ecosystems that can be made up of data extracted directly from the LMS database, data that are generated manually by tutoring coordinators in HEIs, and even surveys applied for concrete measurements such as emotion, engagement, and motivation, to mention a few. It is worth stressing that working with this type of data is a major challenge, as it implies good data quality and scope, as well as privacy and ethics in working with the data. On the other hand, it is also clear that most of the works that present advances in LA issues are from the USA and European countries. Thus, it is necessary to bet on the application of LA in HEIs, to contribute to the institutions by supporting the permanence of students, to teachers, letting them know what is happening in the student's learning process, and to the latter, providing recommendations for them to enjoy their learning process in the best possible way. The challenges faced by learning analytics focused on student behavior are the integration of data sets from diverse environments, advances in technology, and ethical problem solving. Despite these challenges, this research aims at solutions that contribute early on by making recommendations or warnings to achieve student retention.

6. Conclusion

With the SLR on the subject it has been identified that the term LA as we describe it today arises from the year 2011, being the most cited definition the one that arises in the *"1st International Conference on Learning Analytics and Knowledge 2011"*. Therefore, we can say that it is clear at this time to identify that the objective of LA is to improve learning. It should be noted that, being considered a new term, it is often difficult for some to identify the differences with its predecessor, the term Educational Data Mining. It should be noted that although both work with educational data, their approaches are different. Firstly, we have EDM which focuses on helping teachers and students with the analysis of the learning process using popular data mining techniques such as: clustering, association and classification to name a few, we could say that it seeks the transformation of data into relevant information. Unlike LA where the emphasis is on learning outcomes, a better understanding of the learner's behavior and processes, therefore, its objectives are recommendations, predictions, adaptation and personalization, in addition some common methods are usually used such as: classification, clustering and association, in addition to, social network analysis, sentiment analysis, prediction of learner success among others. It should also be noted that the term Tutoring in this context is defined as a process of group or individual accompaniment that a tutor provides to a student during his stay in an IES, with the purpose of contributing to his integral formation, besides influencing the fulfillment of the institutional goals related to the educational quality such as: raising

the terminal efficiency rates and decreasing the failure and desertion rates. On the other hand, in order for LA to achieve this objective, different techniques and methods are used, which are applied to the data offered by the educational platforms. It is important to mention that there are still few studies on this topic, but it has been identified that in Latin America LA has already begun to be used in HEIs. As indicated, it is still a little explored topic with great areas of opportunity. All the studies found present the common denominator of the use of student information regarding their academic history or surveys as instruments for specific measurements such as academic performance, emotions or student motivation during the course; however, there are still minimal cases in which other types of information are used to predict their future behavior. This research opens a gap to resume studies that can integrate academic information that can come from the SIS (Student Information System), LMS tracking data and information from institutional tutoring programs. It should be noted that the complexity of this part will depend on the ease of access to data from these tutoring programs, since there is no specific system or standard for integrating the information, so each institution stores and processes its data in a particular way. The challenge we face with LA is the integration of data ecosystems from different sources, as well as data quality and scope, and data privacy and ethics.

References

[1]. A. Salas Martinez and A. Ramirez Martinell, Application of learning analytics in higher education institutions: A systematic literature review, in 11th International Conference in Software Engineering Research and Innovation (CONISOFT) (2023).

[2]. G. Siemens and R. S. J. d. Baker, Learning analytics and educational data mining: towards communication and collaboration, in Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (2012), p. 252–254.

[3]. D. Gasevic, Y. Tsai, S. Dawson, and A. Pardo, How do we start? an approach to learning analytics adoption in higher education, International Journal of Information and Learning Technology 4, 342 (2019).

[4]. G. Siemens and P. Long, Penetrating the fog: Analytics in learning and education, EDUCAUSE Review 5, 30 (2011).

[5]. S. J. Aguilar, S. A. Karabenick, S. D. Teasley, and C. Baek, Associations between learning analytics dashboard exposure and motivation and self-regulated learning, Computers & Education 162 (2021).

[6]. C. Schumacher and D. Ifenthaler, Features students really expect from learning analytics, Computers in Human Behavior, 397 (2018).

[7]. S. Charleer, A. Moere, J. Klerkx, K. Verbert, and T. D. Laet, Learning analytics dashboards to support adviser-student dialogue, IEEE Transactions on Learning Technologies 11, 389 (2018).

[8]. M. Waheed, K. Kaur, N. Ain, and N. Hussain, Perceived learning outcomes from moodle: An empirical study of intrinsic and extrinsic motivating factors, Information Development 32, 1001 (2016).

[9]. O. Talbi and A. Ouared, Goal-oriented student motivation in learning analytics: How can a requirements-driven approach help? Education and Information Technologies 27 (2022).

[10]. F. S. Kia, S. D. Teasley, M. Hatala, S. A. Karabenick, and M. Kay, How patterns of students dashboard use are related to their achievement and self-regulatory engagement, LAK 2020 Conference Proceedings – Celebrating 10 years of LAK: Shaping the Future of the Field – 10th International Conference on Learning Analytics and Knowledge, 340 (2020).

[11]. I. Jivet, M. Scheffel, H. Drachsler, and M. Specht, Awareness is not enough. pitfalls of learning analytics dashboards in the educational practice, Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning 10474, 12, (2017).

[12]. B. Kitchenham and P. Brereton, A systematic review of systematic review process research in software engineering, Information and Software Technology, 2049 (2013).

[13]. H. Zhang, M. A. Babar, and P. Tell, Identifying relevant studies in software engineering, Information and Software Technology 53 (2011).

[14]. P. O. Silva-Vasquez, V. Y. Rosales-Morales, and E. Benitez-Guerrero, Automatic code generation of user-centered serious games: A decade in review, Program. Comput. Softw. 48, 685–701 (2022).

[15]. S. I. Fernandez Gregorio, L. G. Montane-Jimenez, C. M. Godoy, and V. Y. Rosales-Morales, Architecture for groupware oriented to collaborative medical activities in the rehabilitation of strokes, Program. Comput. Softw. 49, 643–656 (2024).

[16]. J. A. Reyes, The skinny on big data in education: Learning analytics simplified, TechTrends 59, 75 (2015).

[17]. R. Bodily and K. Verbert, Review of Research on Student-Facing Learning Analytics Dashboards and Educational Recommender Systems, IEEE Transactions on Learning Technologies 10, 405 (2017).
[18]. P. Rojas-Castro, Learning Analytics: una revisiA˜de la literatura, EducaciA˜y Educadores 20, 106 (2017).
[19]. D. T. Tempelaar, B. Rienties, and Q. Nguyen, Towards actionable learning analytics using dispositions, IEEE Transactions on Learning Technologies 10, 6 (2017).
[20]. O. Viberg, M. Hatakka, O. B¨alter, and A. Mavroudi, The current landscape of learning analytics in higher education, Computers in Human Behavior 89, 98 (2018).
[21]. C. Herodotou, B. Rienties, A. Boroowa, Z. Zdrahal, and M. Hlosta, A large-scale implementation of predictive learning analytics in higher education: the teachers' role and perspective, Educational Technology Research and Development 67, 1273 (2019).
[22]. J. Guerra, M. Ortiz-Rojas, M. A. Zuniga-Prieto, E. Scheihing, A. Jimenez, T. Broos, T. De Laet, and K. Verbert, Adaptation and evaluation of a learning analytics dashboard to improve academic support at three Latin American universities, British Journal of Educational Technology 51, 973 (2020).
[23]. S. Ranjeeth, T. P. Latchoumi, and P. V. Paul, A Survey on Predictive Models of Learning Analytics, Procedia Computer Science 167, 37 (2020).
[24]. T. De Laet, M. Millecamp, M. Ortiz-Rojas, A. Jimenez, R. Maya, and K. Verbert, Adoption and impact of a learning analytics dashboard supporting the advisor – Student dialogue in a higher education institute in Latin America, British Journal of Educational Technology 51, 1002 (2020).
[25]. C. Guzman-Valenzuela, C. Gomez-Gonzalez, A. Rojas-Murphy Tagle, and A. Lorca-Vyhmeister, Learning analytics in higher education: a preponderance of analytics but very little learning? International Journal of Educational Technology in Higher Education 18, 23 (2021).
[26]. A. Rafique, M. S. Khan, M. H. Jamal, M. Tasadduq, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, Integrating learning analytics and collaborative learning for improving student's academic performance, IEEE Access 9, 167812 (2021).
[27]. C. J. Perez Sanchez, F. Calle-Alonso, and M. A. Vega-Rodrıguez, Learning analytics to predict students' performance: A case study of a neurodidactics-based collaborative learning platform, Education and Information Technologies (2022).
[28]. R. Kaliisa, B. Rienties, A. I. Mørch, and A. Kluge, Social learning analytics in computer-supported collaborative learning environments: A systematic review of empirical studies, Computers and Education Open 3 (2022).
[29]. H. Huang, L. Jew, and D. Qi, Take a MOOC and then drop: A systematic review of mooc engagement pattern and dropout factor, Heliyon 9, e15220 (2023).
[30]. P. Prinsloo, M. Khalil, and S. Slade, Learning analytics as data ecology: a tentative proposal, Journal of Computing in Higher Education, 1 (2023).
[31]. K. Kaur and O. Dahiya, Role of educational data mining and learning analytics techniques used for predictive modeling, in 2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM) (2023) pp. 1–6.
[32]. M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs, A reference model for learning analytics, Int. J. Technol. Enhanc. Learn. 4, 318–331 (2012).
[33]. F. Gutierrez, K. Seipp, X. Ochoa, K. Chiluiza, T. De Laet, and K. Verbert, Lada: A learning analytics dashboard for academic advising, Computers in Human Behavior 107, 105826 (2020).
[34]. J. Kim, I. Jo, and Y. Park, Effects of learning analytics dashboard: Analyzing the relations among dashboard utilization, satisfaction, and learning achievement, Asia Pacific Education Review 17, 13 (2016).

Информация об авторах / Information about authors

Ангел САЛАС-МАРТИНЕС получил степень магистра в области информационных сетей и интегрированных систем в Национальной лаборатории передовых вычислений (LANIA) штата Веракрус (Мексика). В настоящее время обучается в аспирантуре Университета штата Веракрус. Область научных интересов: программная инженерия, учебная аналитика (LA), панель мониторинга аналитики обучения (LAD), интеллектуальный анализ данных (DM), интеллектуальный анализ данных в образовании (EDM).

Angel SALAS-MARTINEZ – holds a Master's degree in Networks and Integrated Systems from National Laboratory of Advanced Computing (LANIA) in Veracruz, Mexico. He is a PhD student

in Computer Science at the Universidad Veracruzana in Mexico. His areas of interest are Software Engineering, Learning Analytics (LA), Learning Analytics Dashboard (LAD), Data Mining (DM) and Educational Data Mining (EDM).

Альберто РАМИРЕС-МАРТИНЕЛЬ имеет степень PhD по исследованиям в области образования, полученную университете Ланкастера (Великобритания); степень магистра по программированию и коммуникациям Университета прикладных наук (Фуртванген, Германия); степень по программной инженерии Национального университета Мексики, степень бакалавра по гуманитарным наукам Университета монастыря Сор-Хуана. В настоящее время работает профессором и исследователем в Университете штата Веракрус (Мексика). Его научные интересы в основном связаны с внедрением информационных и коммуникационных технологий в высших учебных заведениях.

Alberto RAMIREZ-MARTINELL – holds a PhD in Educational Research from Lancaster University, United Kingdom; a Master's degree in Computer Science and Media from the University of Applied Sciences in Furtwangen, Germany; a degree in Computer Engineering from Universidad Nacional Autónoma de México and a BA in Humanities from Universidad del Claustro de Sor Juana. He is currently Full-time Professor and Researcher at the Universidad Veracruzana, Mexico. His main research topic revolves around a disciplinary approach for the incorporation of Information and Communication Technology (ICT) in Higher Education Institutions.

Самюэль МАРТИНЕС-РАМОС – профессор Высшего технологического института в Пероте Национального технологического центра Мексики (TecNM/ITSP). Получил степень магистра в области информационных сетей и интегрированных систем в Национальной лаборатории передовых вычислений (LANIA) Веракруса (Мексика). Сфера научных интересов: разработка технологий, бизнес-аналитика, Интернет вещей.

Samuel MARTINEZ-RAMOS – holds a Master's degree in Networks and Integrated Systems from National Laboratory of Advanced Computing (LANIA) in Veracruz, Mexico. He is currently a Full Time Professor at the Tecnológico Nacional de México / Instituto Tecnológico Superior de Perote. His areas of interest are technology development, Business Intelligence and Internet of Things.

DOI: 10.15514/ISPRAS-2024-36(6)-13

Time Series Models using in Prediction of COVID-19 Infection Cases in Mexico

K. V. Cortés-Martínez, ORCID: 0000-0002-5578-7125 <keila.cortes.mtz@gmail.com>
H. Estrada-Esquivel, ORCID: 0000-0002-1466-7581 <hugo.ee@cenidet.tecnm.mx>
A. Martínez-Rebollar, ORCID: 0000-0002-1071-8599 <alicia.mr@cenidet.tecnm.mx>

TECNM/Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Morelos, México.

Abstract. The COVID-19 pandemic was the first health crisis to affect the entire world in this century. The data captured revealed a lack of organization and control in health measures, containment, and mitigation policies, as well as a lack of planning and coordination in the use of medical supplies, which motivated the development of prediction models that provided predictive information on the evolution of the pandemic. In this work, a time series of accumulated cases of infection was generated through official data provided by the Ministry of Health of the Government of Mexico. Six deterministic and stochastic predictive models were applied to this information to compare their efficiency in predicting cases of COVID-19 infection. These models were applied to data from two cities in Mexico, Colima and the State of Mexico. The study concludes that the ARIMA and ANN MLP models adapt better to the data that is generated daily, therefore, they have an improved prediction capacity.

Keywords: COVID-19; Machine learning; Predictive models; ARIMA; ANN MLP.

For citation: Cortés-Martínez K. V., Estrada-Esquivel H., Martínez-Rebollar A. Time series models using in prediction of COVID-19 infection cases in Mexico. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 231-246. DOI: 10.15514/ISPRAS-2024-36(6)-13.

Прогнозирование случаев заражения COVID-19 в Мексике на основе моделей временных рядов

К. В. Кортес-Мартинес, ORCID: 0000-0002-5578-7125 <keila.cortes.mtz@gmail.com>
Х. Эстрада-Эскивел, ORCID: 0000-0002-1466-7581 <hugo.ee@cenidet.tecnm.mx>
А. Мартинес-Ребояр, ORCID: 0000-0002-1071-8599 <alicia.mr@cenidet.tecnm.mx>
Национальный центр исследований и технологических разработок (TECNM), Куэрнавака, Морелос, Мексика.

Аннотация. Пандемия COVID-19 стала первым кризисом в области здравоохранения, затронувшим весь мир в этом столетии. Собранные данные выявили отсутствие организации и контроля в мерах здравоохранения, сдерживании и смягчении последствий, а также отсутствие планирования и координации в использовании предметов медицинского назначения, что побудило к разработке моделей прогнозирования, которые предоставили прогнозную информацию о развитии пандемии. В этой работе временные ряды накопленных случаев заражения были получены с помощью официальных данных, предоставленных Министерством здравоохранения правительства Мексики. К этой информации были применены шесть детерминированных и стохастических прогностических моделей для сравнения их эффективности в прогнозировании случаев заражения COVID-19. Эти модели были применены к данным из двух городов Мексики, Колимы и штата Мексика. В исследовании делается вывод о том, что модели ARIMA и ANN MLP лучше адаптируются к данным, которые генерируются ежедневно, поэтому они имеют улучшенную способность прогнозирования.

Ключевые слова: пандемия COVID-19; машинное обучение; прогнозирующие модели; модель ARIMA; модель ANN MLP.

Для цитирования: Кортес-Мартинес К. В., Эстрада-Эскивел Х., Мартинес-Ребояр А. Прогнозирование случаев заражения COVID-19 в Мексике на основе моделей временных рядов. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 231–246 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–13.

1. Introduction

The COVID-19 pandemic emerged in December 2019 in the city of Wuhan, China; This event set a precedent in the study of diseases and public health emergency declarations due to its rapid spread globally [1]. In Mexico, the first case was recorded in February 2020 [2], despite containment measures [3], increases in the number of cases were observed, generating multiple waves of contagion until 2021. At this point, various models were created to predict the evolution of the pandemic using time series, but their limitations are still recognized, due to the variability of the data, the appearance of new variants of the virus, other socioeconomic factors, and the effectiveness of various control measures. Therefore, the objective of this paper is to contrast the efficiency of different models used in the literature for the analysis in predicting infection cases, considering only the data collected by the Ministry of Public Health in Mexico [4], through the use of time series. The rest of this work is organized as follows: In section 2, a summary of related works is presented. Section 3 details the methodology proposed in this article, from obtaining and analyzing the data set to the representation and experimentation with the various predictive models found in the literature. Section 4 shows the results and discussion. Finally, section 5 presents the conclusions and future work.

2. Related work on time series prediction

The data that describes the evolution of infections were used in several research groups for implementing predictive models and also to analyze, both statistically [4] and focused on machine learning [5-6], different characteristics of the COVID-19 pandemic, highlighting its effectiveness in different contexts, for example:

- *Prediction of outbreak trends:* Linear prediction models, support vector machines and exponential smoothing were used in the research of [7] and [8].
- *Infection wave prediction:* [9] and [10] applied LSTM and RNN to describe the fluctuations in the increase in the number of infection cases.
- *Prediction of positive cases:* [11] used an MLP to predict the maximum number of positive patients, [12] [13] demonstrated the effectiveness of the ARIMA and Prophet model, [14] [15] used a model with LSTM-GRU to the same task, however this latest work added the calculation of the future transmission of the virus.

In the current debate about which predictive model can efficiently provide, plan and address response strategies and resource allocation, the limitations of each of them are identified, mainly due to the variability and quality of the data that feed the databases, the emergence of new virus variants, the control measures applied and the inclusion of external covariates. Therefore, the objective of this article is to contrast the efficiency of different models used in the literature for the analysis in predicting infection cases, considering only the data collected by the Ministry of Public Health in Mexico [16], through the use of time series.

3. Methodology used in the comparison of prediction models

The methodology was divided into five stages (Fig. 1). In the first, COVID-19 data from Mexico were collected; in the second, they were analyzed and pre-processed to represent them; in the third, they were organized by time series; in the fourth, various prediction models found in the literature were tested. The last stage was the graphing and comparison of the experimentation on the data generated daily during the pandemic.

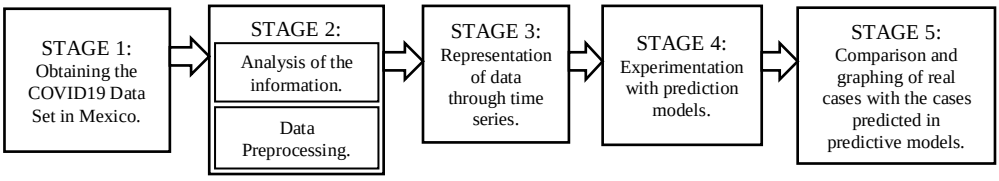


Fig. 1. Methodology for the analysis of time series with prediction models.

3.1 Obtaining the dataset

The data [16] comes from the Ministry of Health of the Government of Mexico, through the General Directorate of Epidemiology, and was reported during the period from April 19th, 2020 to December 31st, 2021. The data contains only the reported cases of COVID-19 in the different Health Centers nationwide distributed in the thirty-two states. The total number of records obtained was 12,698,740, of which 1,986,260 were analyzed, corresponding only to the states of Colima and the State of Mexico. In addition, each of the records can be divided into groups of: infected, deceased, municipality, state, latitude and longitude. Data were fitted to 7- and 30-day time series.

3.2 Analysis and preprocessing of information

Given the nature of the pandemic, the COVID-19 dataset [16] was found to have noise (mis captured information), redundancy, some missing values (mainly in the comorbidity fields) and outdated data, so it was proposed to normalize and transform in such a way that the resulting set was consistent. For the case study of this article, data corresponding to two states in Mexico were selected:

- *Colima:* the state with the smallest population (731,391 inhabitants) [17].
- *State of Mexico:* the state with the largest population (16,992,418 inhabitants) [17].

The state of Colima has a territorial area of 5,625 km^2 while the State of Mexico has 22,500 km^2, that is, 23.23 times larger than Colima.

During the year 2020, in the state of Colima 630,204 possible cases of contagion were reported (Table 1), but only 8,025 were confirmed (Table 2); while in the state of Mexico of the 5,832,576 cases reported (Table 1), only 161,809 infections were confirmed (Table 2), within the Epidemiological Surveillance System for Viral Respiratory Disease [16]. For the year 2021, the number of infection cases in Colima was 33,509 confirmed (Table 2), and for the State of Mexico there were only 427,068 (Table 2).

Table 1. Characteristics of the states of Colima and the State of Mexico [16, 18].

Description	Colima	State of Mexico
Number Inhabitants	731,391	16,992,418
Reported Cases Year 2020	630,204	5,832,576
Reported Cases Year 2021	658,977	11,126,420

Table 2. Data reported from the states of Colima and State of Mexico at the national level [16].

Types of Accumulated Cases	Colima	Estado De México
Confirmed 2020	8,025	161,809
Negatives 2020	7,438	210,710
Suspects 2020	1,802	56,530
Deaths 2020	820	23,961
Confirmed 2021	33,509	427,068
Negatived 2021	44,965	864,552
Suspects 2021	2,554	95,033
Deaths 2021	2,041	45,443

According to the number of cases registered at the national level (Table 1), Colima reported until 2021 almost 90% of cases with respect to its total population, and the State of Mexico up to 65%. A possible cause for the reporting of a greater number of cases in the state of Colima may be related to population density and its territorial distribution, since Colima has 130 inhabitants/km^2 on average, while the state of Mexico has 760. inhabitants/km^2, the latter concentrating the largest population in its capital with respect to its 125 municipalities.

Once the data was reviewed and normalized, preprocessing was done, for this a Univariate analysis was carried out in R and Python, to understand the distribution, central tendency, dispersion, and other aspects of each variable in the data set, through a statistical summary and application of Filter Methods such as ANOVA and Chi-Square [19]. In this stage, the objective was that each of the characteristics obtained in the data set were within the same scale (for example, in the sex attribute, the categorical values were 1 = woman, 2 = man and they jumped up to category 99 = not specified), with the purpose that the values of its distribution and frequency did not affect the interpretation of the predictive model. Some values were also imputed, such as the age attribute, where sometimes atypical data were found but since they were scarce, values were assigned using the median. Subsequently, SHAP (SHapley Additive exPlanations) [20] was used, because SHAP considers all possible combinations of features and calculates the average marginal contribution of each feature across these combinations, thus identifying which attribute is most important in a prediction and detecting biases or inconsistencies in the predictive model. For our case study, this tool showed that some attributes were irrelevant (such as those referring to comorbidities), helping to reduce the dimensionality of our dataset. Finally, we verified the homoscedasticity of the data with the Box-Cox Transformation. For future experiments, the accumulated cases were counted to obtain the number of confirmed cases of infection, negative cases, suspected cases, deaths, and recovered cases.

3.3 Representation of data using time series

A crucial tool to understand, manage and control the spread of a disease is the time series [21]; to represent the information from the states of Colima and Mexico, the total number of infection cases reported during the years 2020 and 2021 was counted by daily date and by week (Fig. 2 and 3).

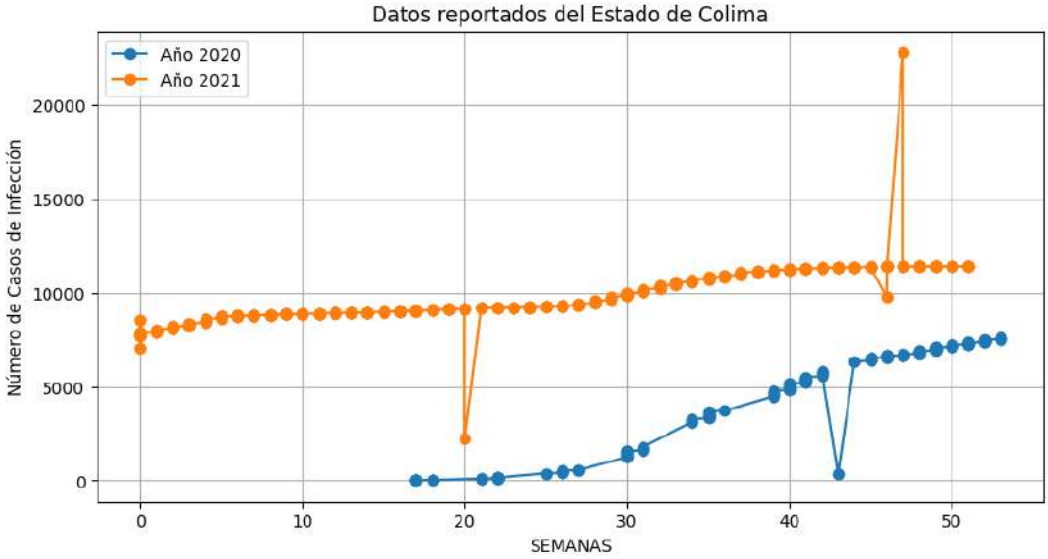


Fig. 2. Time series of accumulated cases of COVID-19 infection in the state of Colima, period April 2020 to December 2021.

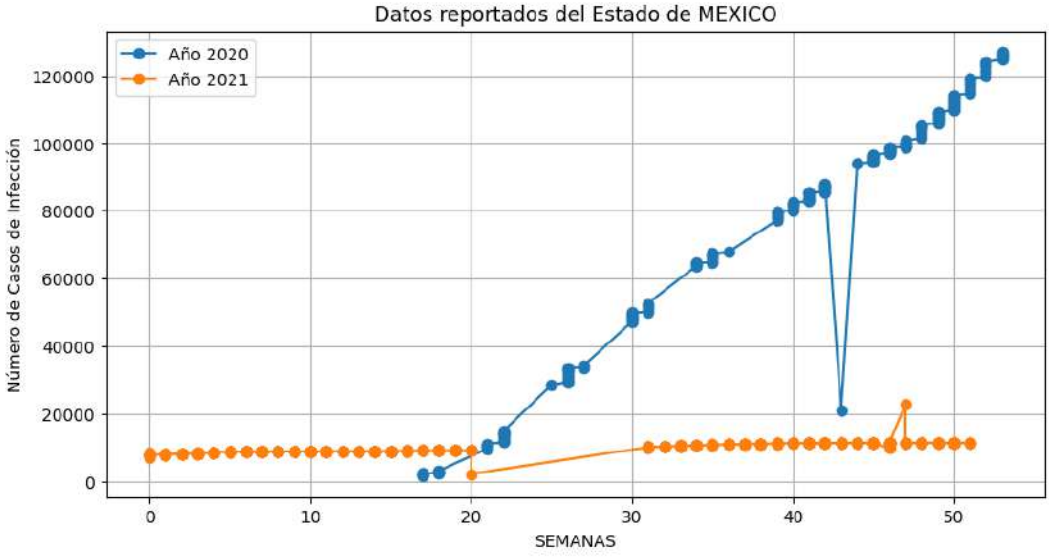


Fig. 3. Time series of accumulated cases of COVID-19 infection in the State of Mexico, period April 2020 to December 2021.

In the two previous series, the similarity can be seen in week 42 of the year 2020 (January 4th, 2021 and January 6th, 2021) and week 20 of the year 2021 (January 4th, 2021 and January 7th, 2021) where a downward peak is observed, this is explained because changes were made to the records, and new attributes were added: PCR test information and new classifications. In week 45 of 2021 year, Colima registered a decrease due to the cases of deaths reported compared to the cases of infection;

In week 46 of 2021 year (corresponding to the dates from November 15th, 2021 to November 21st, 2021), both states had an uptick in contagion cases. The above could have derived from the case of omission in the containment measures with respect to the previous two weeks, since in Mexico the month of November is a month of traditions, in addition to the fact that approximately 80% of the population was already vaccinated with the first dose [22]. In the distribution of the data for both states, a certain seasonal component can be understood, because a stable range without trend is maintained in the periods of growth and decrease. Given the complexity of the problem addressed, in terms of the amount of information collected, the trends were exemplified on an annual, monthly, and weekly basis, taking periods of 7 days and 30 days for experimentation.

4. Results in the comparison of prediction models

4.1 Experimentation with prediction models

In the literature, various models have been found that address the problem of estimating and/or predicting the number of infection cases through time series [23] [24]. Therefore, six models were experimented with that used different criteria, concepts, and methodologies, classifying them as deterministic and stochastic (Fig. 4). In the deterministic model [25] the values in its parameters are usually constant and rigid (controlled behavior), while in the stochastic model [25] the parameters used have random or estimated values (probabilistic behavior), that is, according to the literature, the latter captures the behavior of diseases more realistically. Once the data and predictive models were identified, they were compared and experimented with each of them.

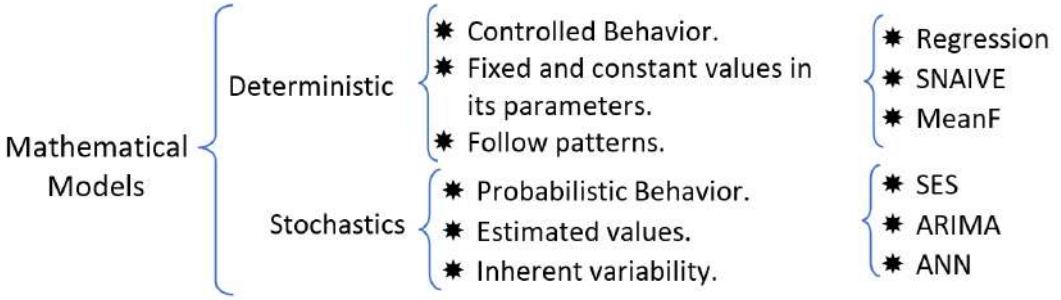


Fig. 4. Classification of mathematical models in the prediction of COVID-19 infection cases.

4.2 Comparison between Regression, SNAIVE and MeanF

The regression model [26-27] allows the incorporation of various predictor variables, however, it attempts to place the value of its parameters in a linear manner, which makes its implementation difficult to represent the real world, that is, collinearity in it can lead to imprecise estimates, so their confidence intervals are wide. The SNaive (Simple Smoothing) [28] and MeanF (Averaging Method) [29-30] models are methods based on past values of a time series, so their prediction focuses on sequential data, so they do not adequately capture the trend. seasonality and randomness. noise. These models are represented with the following equations (Table 3).

As seen in Fig. 5, using the Regression model [27], the prediction does not converge with almost any data, however, it seems to exemplify the trend of the data. In Fig. 6, for the State of Mexico, the regression model simply did not reflect a prediction during the year 2020 with an observed period of 7 days; and for the data for the year 2021, it only reflected the trend they had.

Since the SNaive model [28] focuses on the estimated value of the average of the previous values, in the graphs of both the state of Colima (Fig. 7) and the state of Mexico (Fig. 8) similarities were observed in the behavior of the real values with the predicted ones, although the latter are far from the real data.

Table 3. Equations of Regression, SNaive y MeanF models.

Model	Equation		Parameters
Regression	$Y = \beta_0 + \beta_1 X + \varepsilon$	(5)	Y is dependent variable (that wants it to predict). β_0 is the ordinate at the origin or intercept. β_1 is the slope of a regression line. X is the independent variable (used to make the prediction). ε is the error that represents the variability not explained by the model.
SNAIVE	$\check{Y}_{t+1} = Y_t$	(6)	$\check{Y}$ represents the prediction for time $t+1$. Y_t is the last value observed in time. t time.
MeanF	$\check{Y}_{t+1} = \frac{1}{t}\sum_{i=1}^{t} Y_i$	(7)	$\check{Y}_{t+1}$ represents the prediction for time $t+1$. Y_i are values observed at the previous times, from $i = 1$ to $i = t$. t is the number of past observations in the time series.

The MeanF model [30], as its name indicates, is based on calculating the arithmetic mean of past values in a time series to predict the future value, so it follows a constant pattern based on historical observations, that is, its relationship with time series lies in making forecasts on sequential data. For the infection cases from the state of Colima (Fig. 9), given that the data was scarce, only when these were described every 7 days was it possible to capture a pattern similar to that of the real data, however, in obtaining predictive data at 30 days, this model did not even exemplify a trend. With the experimentation of the MeanF model in the state of Mexico (Fig. 10), during 2020 year the data observed for 7 days had wider fluctuations compared to the real data; for the information collected in 2021, MeanF became more robust and less sensitive to changes, reducing errors between observed and predicted data.

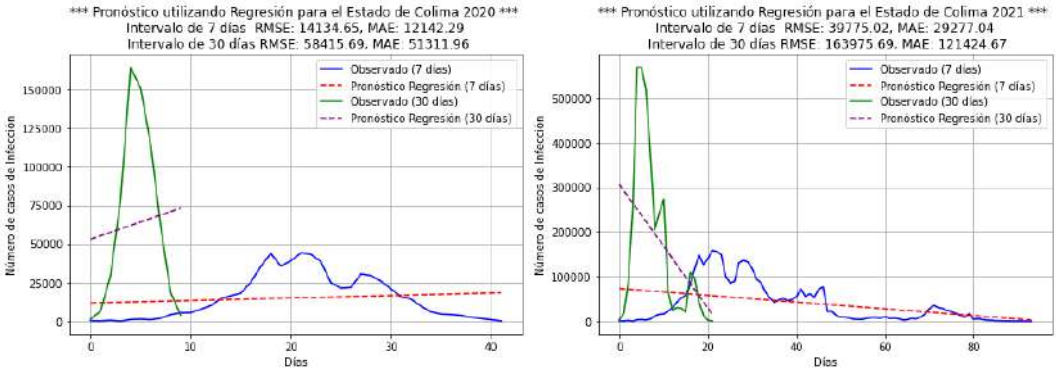


Fig. 5. Regression Model of the state of Colima.

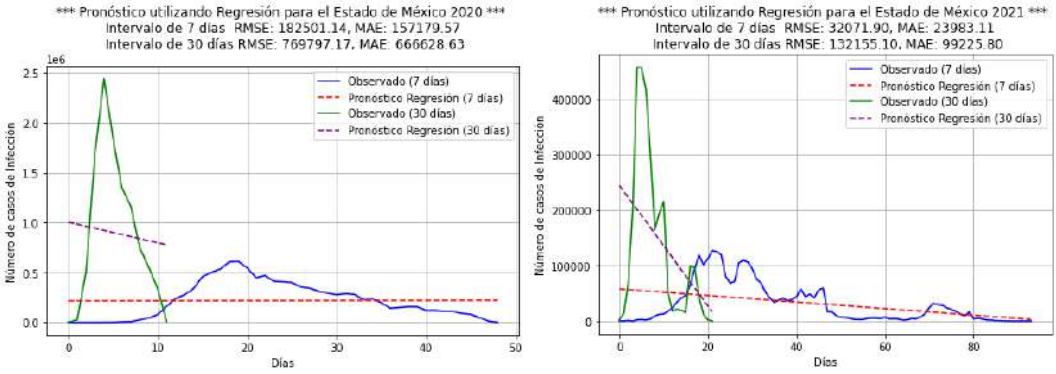


Fig. 6. Regression Model of the state of Mexico.

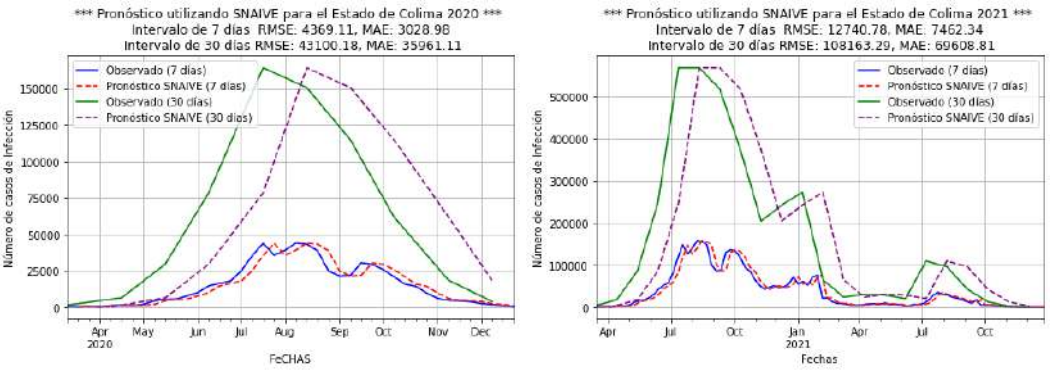


Fig. 7. SNAIVE model of the state of Colima.

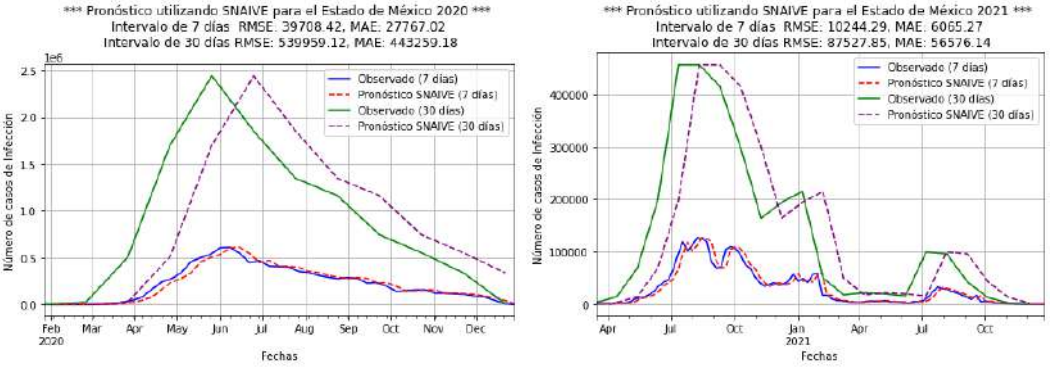


Fig. 8. SNAIVE model of the state of México.

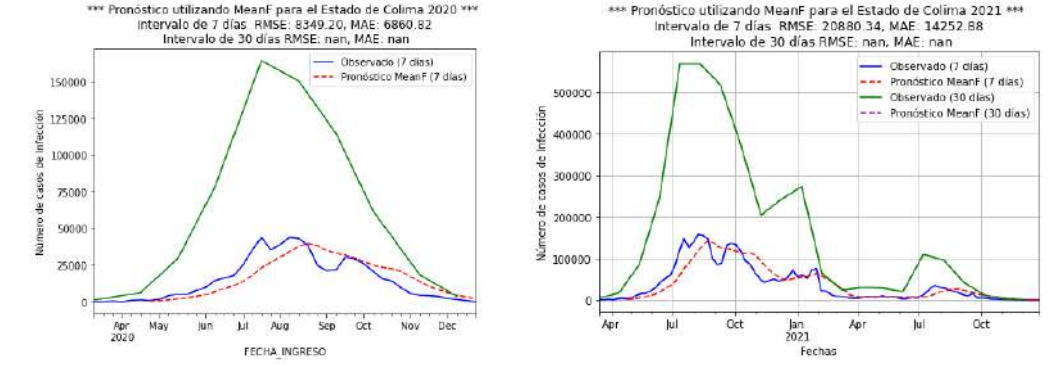


Fig. 9. MeanF model of the state of Colima.

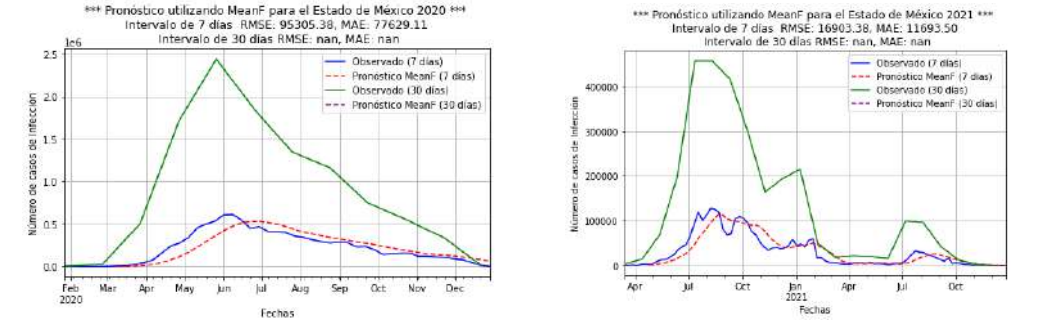


Fig. 10. MeanF model of the state of México.

4.3 Comparison between SES, ARIMA, ANN

The SES [29], ARIMA [30] and ANN [31-32] models are of the stochastic type and explicitly capture the variability of past information, that is, they estimate their predictions from a tuning of the errors generated from historical data, which generates greater certainty. These models are represented by the equations shown in Table 4.

Table 4. Equations of the SES, ARIMA y ANN models.

Model	Equation	Parameters
SES	$\check{Y}_{t+1} = \alpha \cdot Y_t + (1-\alpha) \cdot \check{Y}_t$ (8)	$\check{Y}_{t+1}$ is the prediction for time $t+1$. Y_t is the value observed at time t. $\check{Y}_t$ is the prediction for time t (the predicted value in the previous period). α is the smoothing factor, a value between 0 and 1 that controls the weight of the most recent observation to the previous prediction.
ARIMA	$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$ (9)	Y_t is the value of the time series at time t. c is a constant $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregression parameters, which represent the relationship between the current value and the past values of the series. $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average parameters, which represent the relationship between the current value and the past errors of the series. ε_t is the error term at time t, reflecting variability not explained by the autoregressive and moving average components of the model.
ANN	$z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}\ a^{[l]} = g(z^{[l]})$ (10)	$z^{[l]}$ is the activation vector before applying the activation function at layer l. $W^{[l]}$ is the weight matrix associated with layer l. $a^{[l-1]}$ is the activation vector of the previous layer. $b^{[l]}$ is the bias vector at layer l. $g()$ is the activation function applied element by element.

Although the SES [29] model is easy to implement compared to others such as ANN [31, 33] and ARIMA [30, 32], in this experimentation, one of its limitations in prediction was the need to assign an alpha value, which controls the most recent observation, which leads to a lack of long-term memory, consequently it falls into an analysis where the trend of the data must be more linear or constant, no matter if the amount of information is little (Fig. 11) or a lot (Fig. 12) to predict. In this experimentation, the SES model simply did not converge with any of the data.

The ARIMA (Autoregressive Integrated Moving Average) model [31] consists of 3 main components:

- AR (Autoregressive): which models the dependence of a current observation on past observations.
- I (Integrated): which indicates the stationarity of a time series.
- MA (Moving Average): indicates the relationship of the dependence between a current observation and past errors, using moving average coefficients.

In the ARIMA experiment, the time series of the state of Colima (Fig. 13) and the State of Mexico (Fig. 14) presented the lowest error, likewise, the prediction reflected a more adequate behavior concerning the real data.

Finally, an ANN MLP (Multi-Layer Perceptron) model [32] was experimented with, which, by its nature, can solve problems that are not linearly separable in addition to capturing complex relationships between variables. In this experiment, for the case of the state of Colima (Fig. 15), the predicted data matched the real data better when they were studied for 7 days.

With the data from the state of Mexico (Fig. 16), a better result will also be considered when these are analyzed in a shorter time.

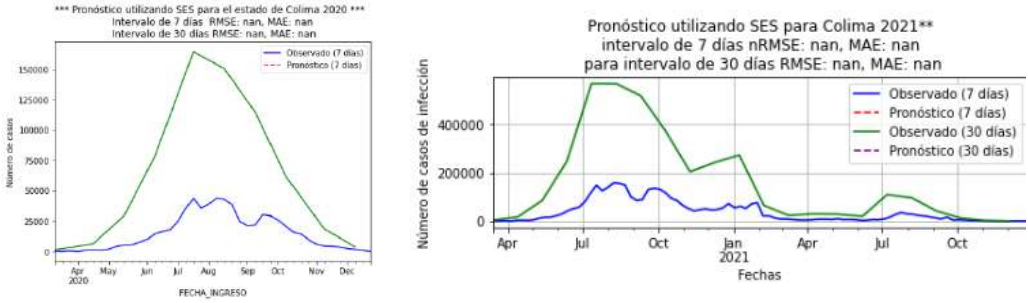


Fig. 11. SES model of the state of Colima.

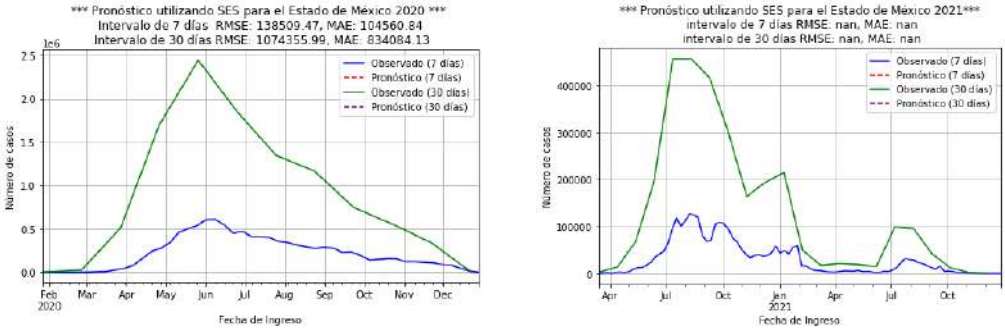


Fig. 12. SES model of the state of México.

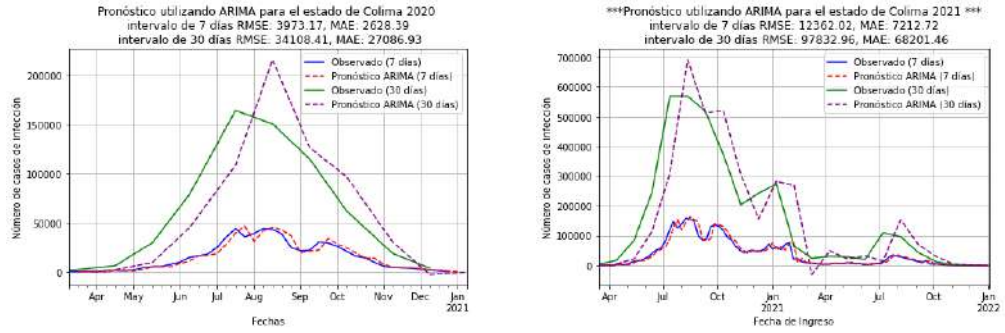


Fig. 13. ARIMA model of the state of Colima.

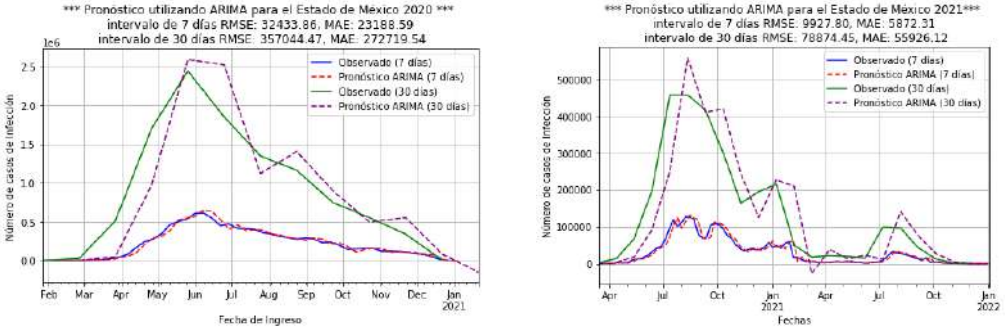


Fig. 14. ARIMA model of the state of México.

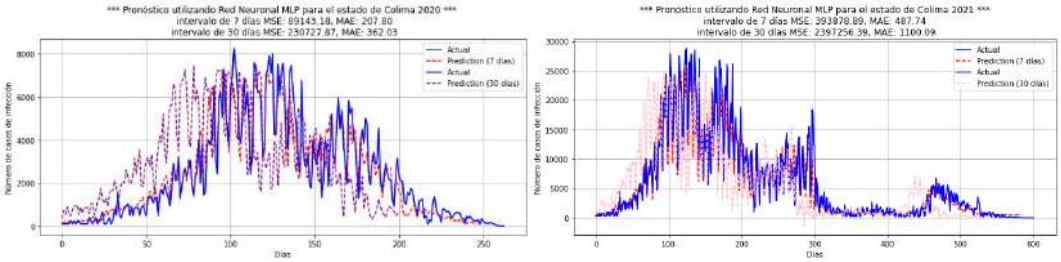


Fig. 15. MLP model of the state of Colima.

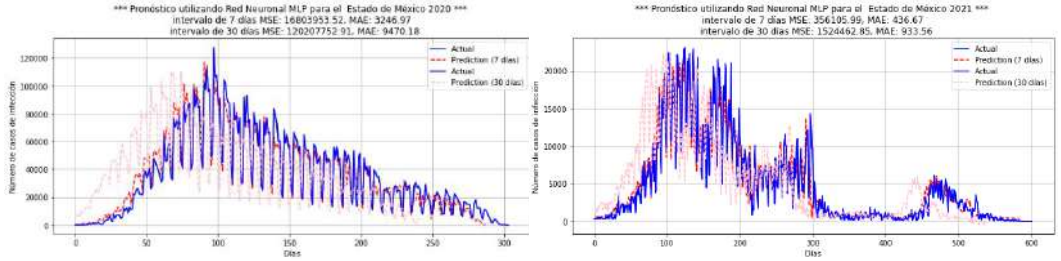


Fig. 16. MLP model of the state of México.

4.4 Discussion of results

To evaluate each of the predictive models, the averaged error biases were obtained (from the real data with the predicted data), and they were evaluated with the RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) metrics. A low RMSE indicates that the model is producing accurate predictions, however, because errors are first squared before averaging, this metric comes to penalize large errors more. Likewise, if an MAE is low, it indicates that the model has accurate predictions, which, due to its nature, only shows the difference between the predicted value and the actual value. The following Tables (Tables 5 and 6) show the data resulting from the application of these metrics.

Table 5. Description of the predictive models in the time series of the State of Colima.

State | Year	COLIMA 2020				COLIMA 2021			
Period	7 days		30 days		7 days		30 days	
Model | Error	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Regression	14,134.65	*12,142.29*	58,415.69	*51,311.96*	39,775.02	*29,277.04*	163,975.69	*121,424.67*
SNAIVE	4,369.11	3,028.98	43,100.18	35,961.11	12,740.78	7,462.34	108,163.29	69,608.81
MeanF	8,349.20	6,860.82	NAN	NAN	20,880.34	14,252.88	NAN	NAN
SES	NAN	NAN	NAN	NAN	NAN	NAN	NAN	NAN
ARIMA	3,973.17	2,628.39	34,108.41	27,086.93	12,362.02	7,212.72	97,832.96	68,201.46
MLP	*89,143.18*	**207.80**	*230,727.87*	**362.03**	*393,878.89*	**487.74**	*2,397,256.39*	**1,100.09**

**Notation: Numbers in italics indicate high error values. Shaded values indicate low values. Values in bold indicate the lowest values. NAN notes that the model did not make the prediction.*

Table 6. Description of the predictive models in the time series of the State of Mexico.

State | Year	STATE OF MÉXICO 2020				STATE OF MÉXICO 2021			
Period	7 days		30 days		7 days		30 days	
Model | Error	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Regression	182,501.14	*157,179.57*	769,797.17	666,628.63	32,071.90	*23,983.11*	132,155.10	*99,225.80*
SNAIVE	39,708.42	27,767.02	539,959.12	443,259.18	10,244.29	6,065.27	87,527.85	56,576.14
MeanF	95,305.38	77,629.11	NAN	NAN	16,903.38	11,693.50	NAN	NAN
SES	138,509.47	104,560.84	1,074,355.99	*834,084.13*	NAN	NAN	NAN	NAN
ARIMA	32,433.86	23,188.59	357,044.47	272,719.54	9,927.80	5,872.31	78,874.45	55,926.12
MLP	*16,803,953.52*	**3,246.97**	*120207752.91*	9,470.18	*356,105.99*	**436.67**	*1,524,462.85*	**933.56**

**Notation: Numbers in italics indicate high error values. Shaded values indicate low values. Values in bold indicate the lowest values. NAN notes that the model did not make the prediction.*

Regarding the period analyzed, the lowest errors occurred in the data analyzed for 7 days. The individual results of the MeanF and SES models showed that they were not able to predict data at some point in the experimentation, because they require controlling or averaging previous data so that the prediction has a value between 0 and 1. The results of the Regression model frequently presented the highest error values in the MAE metric, which indicates that the predicted data never converged with the real data.

The SNAIVE model, although it presented errors, the predicted data in its graph showed behavior similar to the real data. The values with the lowest errors, shown in Tables 5 and 6, concerning the experimental models were those of ARIMA and MLP (shaded in gray); Although the ANN MLP came to present the highest values of errors in the RMSE metric, this only tells us that it usually adjusts/adapts better to a longer period of time in addition to reducing the error more significantly at a greater number. of data. The lowest results were presented by the MLP model (shaded in Bold), with an observed period of 7 days. Within the previous graphs (section 3), you can see more clearly the projection of the real data and the predicted data.

5. Conclusions and future work

Taking into consideration the set of data provided by the Ministry of Health of Mexico, some of the predictive models did not show the expected predictive behavior; however, the results shown by ARIMA and ANN showed lower errors compared to the other models. The prediction analysis carried out in this work contrasts with the results of other investigations; unfortunately, most of these do not share the parameters used, and the results usually vary concerning the data set they study in their experimentation, which makes it difficult to improve the models and/or be used as a reference to continue with new research. With the results obtained in this research, it is proposed to develop a model that focuses on prediction, not only of cases of infection of the COVID-19 pandemic but also of other diseases, adding other factors and/or experimenting with other types of models of ANN like the LSTM.

Code Availability

All results for this study were generated using R and Python. The code that generated the figures in this article is available at https://github.com/KeilaVCortes/COVID19_prediction.

References

[1]. Organización Mundial de la Salud (OMS), “Enfermedad por el Coronavirus,” Novel Coronavirus 2019, 42–49 (2021). [Last access: February 28th, 2021]. https://www.paho.org/es/enfermedad-por-coronavirus-covid-19.

[2]. Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI), “Línea de tiempo COVID-19 en México”, [Last access: December 27th,2023]. https://micrositios.inai.org.mx/conferenciascovid-19tp/?page_id=8432.

[3]. Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI), “Sugerencias de Actividades para realizar durante la contingencia sanitaria #Quédate en Casa”, pp. 1-45 (2021). https://inicio.inai.org.mx/nuevo/RecomendacionesINAICOVID-19.pdf.

[4]. J. C. Clement, V. K. Ponnusamy, K. Sriharipriya, et al., “A Survey on Mathematical, Machine Learning and Deep Learning Models for COVID-19 Transmission and Diagnosis”, IEEE reviews in biomedical engineering, 15, 325–340(2022). https://doi.org/10.1109/RBME.2021.3069213.

[5]. A. Mehrotraa y R. Agarwal, “A Review of Use of Data Mining duringCOVID-19 Pandemic”, Turkish Journal of Computer and Mathematics Education, 12(6), 4547-4552(2021). https://doi.org/10.17762/turcomat.v12i6.8441.

[6]. Vasilev, I.A., Petrovskiy, M.I., Mashechkin, I.V, et al. “Predicting COVID-19-Induced Lung Damage Based on Machine Learning Methods”. Programming and Computer Software, 48, 243–255 (2022).doi: https://doi.org/10.1134/S0361768822040065.
[7]. F. Rustam, A., et al. “COVID-19 Future Forecasting Using Supervised Machine Learning Models”, IEEE Access Xplore, 8, 101489-101499(2020). doi:10.1109/ACCESS.2020.2997311.
[8]. Madini O. Alassafi, Mutasem Jarrah, Reem Alotaibi, “Time series predicting of COVID-19 based on deep learning”, Neurocomputing 468, 335-344 (2022). ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.10.035.
[9]. R. Chandra, A. Jain y D. Singh Chauhan, “Deep learning via LSTM models for COVID-19 infection forecasting in India”, PLOS ONE, 17(1), 1-28 (2022), https://doi.org/10.1371/journal.pone.0262708.
[10]. S. Kumar, R. Sharma, T. Tsunoda, et al., “Forecasting the spread of COVID-19 using LSTM network”. BMC Bioinformatics, 22 (6), 316 (2021). https://doi.org/10.1186/s12859-021-04224-2.
[11]. Z. Car, S. Baressi Šegota, N. Anđelić, et al., “Modeling the Spread of COVID-19 Infection Using a Multilayer Perceptron”, Hindawi. Computational and Mathematical Methods in Medicine, 2020 (5714714), 1-10 (2020). https://doi.org/10.1155/2020/5714714.
[12]. R. Somyanonthanakul, K. Warin, W. Amasiri, et al., “Forecasting COVID-19 cases using time series modeling and association rule mining”, BMC Medical Research Methodology, 22(281), 1-18 (2022), doi: 10.1186/s12874-022-01755-x.
[13]. M. Abdelaziz, A. Ahmed, A. Riad, et al., “Forecasting daily confirmed COVID-19 cases in Algeria using ARIMA models”, Short research communication- EMHJ, 29 (7), 515-519 (2023), https://doi.org/10.26719/emhj.23.054.
[14]. S. Sah, B. Surendiran, R. Dhanalakshmi, et al., “Forecasting COVID-19 Pandemic Using Prophet, ARIMA, and Hybrid Stacked LSTM-GRU Models in India”, Hindawi. Computational and Mathematical Methods in Medicine, 2022 (1556025), 1-19 (2022), https://doi.org/10.1155/2022/1556025.
[15]. Y. Jin, R. Wang, X. Zhuang, et al., “Prediction of COVID-19 Data Using an ARIMA-LSTM Hybrid Forecast Model”, MDPI Mathematics, 10 (21), 1-13 (2022), https://doi.org/10.3390/math10214001.
[16]. Dirección General de Epidemiología, “Datos Abiertos Bases Históricas”. Available at.: https://www.gob.mx/salud/documentos/datos-abiertos-bases-historicas-direccion-general-de-epidemiologia, [Last access: January 25th, 2024].
[17]. Instituto Nacional de Estadística y Geografía (INEGI), “Consulta de indicadores sociodemográficos y económicos por área geográfica”. Available at.: https://www.inegi.org.mx/, [Last access: Decembre 27th,2023].
[18]. Instituto Nacional de Estadística y Geografía (INEGI) para niños. “Densidad de Población”, Available at.: https://cuentame.inegi.org.mx/poblacion/densidad.aspx?tema=P, [Last access: Decembre 27th,2023].
[19]. G. Ramírez-Valverde y B. Ramírez-Valverde, “Modelo estadístico para defunciones y casos positivos de COVID-19 en México”, EconoQuantum, 18(1), 1-20 (2021). https://doi.org/10.18381/eq.v18i1.7223.
[20]. Rojas-García, M., Vázquez, B., Torres-Poveda, K.et al. “Lethality risk markers by sex and age-group for COVID-19 in Mexico: a cross-sectional study based on machine learning approach”. BMC Infectious Diseases, 23(18), 1-14 (2023). https://doi.org/10.1186/s12879-022-07951-w.
[21]. Cueva Moncayo, M., Jami Carrera, J., & Betancourt Rubio, E., “Previsión de COVID-19 mediante modelo de series temporales en pacientes de una clínica de salud rural”, Revista Cubana de Investigaciones Biomédicas, 42(3127), 1-21 (2023), Editorial de Ciencias Médicas (ECIMED). https://revibiomedica.sld.cu/index.php/ibi/article/view/3127.
[22]. Secretaría del Gobierno de México (gob.mx), “Boletín estadístico 2021 del exceso de mortalidad por todas las causas durante la emergencia por COVID-19”, Grupo Interinstitucional para la estimación del exceso de mortalidad por todas las causas, 46 (27), 1-42 (2021). https://coronavirus.gob.mx/wp-content/uploads/2021/09/2021.09.08-Boleti%CC%81n_XXI_Exceso_Mortalidad_SE33_MX.pdf.
[23]. Z. Malki, E.-S. Atlam, A. Ewis, et al., “The COVID-19 pandemic: prediction study based on machine learning models”, Environmental Science and Pollution, 28(30), 40496-40506 (2021). doi: 10.1007/s11356-021-13824-7.
[24]. L. Tomov, L. Chervenkov, D. G. Miteva, et al. “Applications of time series analysis in epidemiology: Literature review and our experience during COVID-19 pandemic”, WJCC (World Journal of Clinical Cases), 11 (29), 6974-6983 (2023). PMID: 37946767 doi: 10.12998/wjcc.v11.i29.6974.
[25]. M. S. Bartlett, “Deterministic and Stochastic Models for Recurrent Epidemics”, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, IV, 81-109 (1956). https://apps.dtic.mil/sti/pdfs/AD1028592.pdf.

[26]. A. M. Radwan, “Forecasting of Covid-19 Using Time Series Regression Models”, 2021 Palestinian International Conference on Information and Communication Technology (PICICT), 7-12 (2021). doi: 10.1109/PICICT53635.2021.00014.

[27]. S. Bhutia, B. Patra y M. Ray, “COVID-19 epidemic: analysis and prediction”, IAES International Journal of Artificial Intelligence (IJ-AI), 11(2), 736-745 (2022). ISSN: 2252-8938, doi: 10.11591/ijai.v11.i2.

[28]. V. Chaurasia y S. Pal, “Application of machine learning time series analysis for prediction COVID-19 pandemic”, Research on Biomedical Engineering. Sociedade Brasileira de Engenharia Biomedica, 38(1), 35-47 (2020) . PMCID: PMC7585491. doi: 10.1007/s42600-020-00105-4.

[29]. Fadhil A. Abidi; Zainb Hasan Radiy, “Using exponential smoothing methods to analysis COVID-19 time series”, 2nd International Conference on Engineering and Science to Achieve the Sustainable Development Goals. AIP Conference Proceedings, 3092 (1), id.070014, 1-10 (2024), https://doi.org/10.1063/5.0200417.

[30]. G. Sanglier Contreras, M. Robas Mora, P. Jimenez Gomez, "Use of quantitative forecasting methods and error calculation for better adaptability to the application of a mathematical model to determine the speed of spread of a coronavirus infection (COVID-19) in Spain", Contemporary Engineering Sciences, 13(1), 157-175 (2020), doi: 10.12988/ces.2020.91570.

[31]. R. Somyanonthanakul, K. Warin, W. Amasiri, et al. “Forecasting COVID-19 cases using time series modeling and association rule mining”, BMC Medical Research Methodoly, 22 (281), 1-18 (2022), https://doi.org/10.1186/s12874-022-01755-x.

[32]. Yu, C. S., Chang, S. S., Chang, et al., “A COVID-19 Pandemic Artificial Intelligence-Based System with Deep Learning Forecasting and Automatic Statistical Data Acquisition: Development and Implementation Study”, Journal of medical Internet research, 23(5), (2021). e27806. https://doi.org/10.2196/27806.

[33]. Kadirliev T.M., Telpukhov D.V., Solovyev R.A. Application of Neural Networks for Image Segmentation in the Problem of Fast Global Routing. Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS). 2023;35(5):145-156. (In Russ.) https://doi.org/10.15514/ISPRAS-2022-35(5)- 10.

Информация об авторах / Information about authors

Кейла Басти КОРТЕС-МАРТИНЕС в 2019 году получила степень магистра компьютерных наук в Национальном центре исследований и технологических разработок в Куэрнаваке (Мексика). Ее научные интересы – машинное обучение, распознавание образов и интеллектуальный анализ данных.

Keila Vasthi CORTÉS-MARTÍNEZ received a master’s degree in computer science from the National Center for Research and Technology Development of Cuernavaca, México, in 2019. Her research interests are machine learning, image recognition and data mining.

Хуго ЭСТРАДА-ЭСКИВЕЛ получил степень магистра в области компьютерных наук в Национальном центре исследований и технологических разработок в Куэрнаваке (Мексика) и степень PhD в области информатики в Техническом университете Валенсии (Испания), а также степень PhD в области информатики и телекоммуникаций в Университете Тренто (Италия) в 2008 году. Научный сотрудник исследовательского центра INFOTEC, Национального совета по науке и технологиям и в настоящее время в Национальном центре исследований и технологических разработок CENIDET. Его научные интересы включают большие данные, интернет вещей и умные города.

Hugo ESTRADA-ESQUIVEL received a master’s degree in Computer Science from the National Center for Research and Technology Development of Cuernavaca, México, and a PhD degree in Informatics from the Technical University of Valencia, Spain, and a Ph.D. degree in Informatics and Telecommunications from the Trento University Italy, in 2008. PhD Estrada has been researcher in INFOTEC research center, the National Council of Science and Technology and currently in the National Center for Research and Technological Development, CENIDET. His research interests include big data, Internet of things and smart cities.

Алисия МАРТИНЕС-РЕБОЯР получила степень магистра в области компьютерных наук в Национальном центре исследований и технологических разработок в Куэрнаваке (Мексика) и степень PhD в области информатики в Техническом университете Валенсии (Испания), а также степень доктора философии в области информатики и телекоммуникаций в университете Тренто (Италия) в 2008 году. С 2009 года она является профессором-исследователем Национального центра исследований и технологических разработок CENIDET. Автор 4 книг и более 130 работ. Ее научные интересы включают большие данные, Интернет вещей, умные города и аффективные вычисления.

Alicia MARTINEZ-REBOLLAR received her master degree in Computer Science from the National Center for Research and Technology Development of Cuernavaca, México, and a PhD degree in Informatics from the Technical University of Valencia, Spain, and a PhD degree in Informatics and Telecommunications from the Trento University Italy, in 2008. Since 2009, she has been Research professor with the National Center for Research and Technological Development, CENIDET. She is the author of 4 books, and more than 130 papers. Her research interests include big data, Internet of things, smart cities, and affective computing.