

Предисловие

В 25-м томе Трудов Института системного программирования РАН публикуются статьи, написанные сотрудниками ИСП РАН и других организаций. В этих статьях описываются результаты исследований, выполненных во второй половине 2013 г.

Пять из двенадцати статей 25-го тома посвящены разным аспектам анализа исходного кода программ. В статье С.П. Вартанова и А.Ю. Герасимова «Применение динамического анализа для поиска дефектов в программах на языке Java» рассматриваются методы анализа программ и описывается применение этих методов для автоматизации задачи поиска ошибок в программном обеспечении. Подробно обсуждается метод динамического анализа программ на основе инструментации, отслеживания потока помеченных данных и генерации наборов условий для автоматического построения входных данных. Описывается прототип инструмента, реализующего такой подход, представлены результаты его применения для анализа набора программ на языке Java.

Статья М.К. Ермакова и А.Ю. Герасимова «Avalanche: применение параллельного и распределенного динамического анализа программ для ускорения поиска дефектов и уязвимостей» посвящена подходу к уменьшению времени динамического анализа программ за счет применения параллельных и распределенных вычислений при проверке выполнимости ограничений, а также в процессе динамического анализа программ. Приводятся результаты практического применения данного подхода, реализованного в рамках инструмента динамического анализа Avalanche.

Н.Г. Зельцер в статье «Поиск повторяющихся фрагментов исходного кода при автоматическом рефакторинге» обсуждает возможность совмещения автоматического рефакторинга с обнаружением повторяющихся фрагментов исходного кода для программ на языках C/C++. Предлагается классификация программных клонов с точки зрения дальнейшего применения к ним автоматического рефакторинга. Для каждого выделенного типа клонов описывается способ их поиска. Анализируются недостатки существующих инструментов и демонстрируется, что предложенные методы в соответствующих случаях работают корректно.

В статье Н.Л. Луговского и С.В. Сыромятникова «Применение языка KAST для преобразования исходного кода и автоматического исправления дефектов» описывается расширение языка KAST для решения задачи

трансформации исходного кода. Обсуждаются некоторые существующие подходы к трансформации исходного кода и показываются преимущества использования языка KAST.

Статья «Подход к обнаружению ошибок несоответствия типов в коде на динамических языках программирования» И.Е. Бронштейна посвящена обнаружению дефектов в программах, написанных на динамических языках программирования. Демонстрируется, что большинство существующих средств не в состоянии обнаруживать целый класс дефектов: ошибки несоответствия типов. Предлагается подход к выводу типов для динамических языков, а также к реализации обнаружителей дефектов на его основе.

К тематике тестирования программных средств на основе использования формальных моделей относится статья И.С. Захарова и др. «Моделирование окружения драйверов устройств операционной системы Linux». В статье отмечается, что верификация драйвера в комбинации с исходным кодом сердцевины ядра операционной системы не представляется возможной из-за сложности и объема получаемого кода. В качестве решения этой проблемы предлагаются методы моделирования окружения драйверов на основе π исчисления Р.Милнера и трансляции π модели окружения в программу на языке Си, которая при связывании с исходным кодом драйвера описывает с точки зрения инструментов статической верификации те же сценарии работы драйвера, что и реальное окружение драйвера в операционной системе.

Следующие четыре статьи затрагивают тематику управления данными и их анализа. В статье С.Д. Кузнецова и Н.А. Мендковича «Оценка эффективности минимизации ограничений запросов к СУБД» описываются усовершенствованные алгоритмы лексической оптимизации запросов. Эти алгоритмы обнаруживают избыточные элементы в условии выборки запроса и удаляют их. Представлены результаты применения этих оптимизационных методов и их влияние на скорость обработки запросов.

Статья В.А. Золотова и В.А. Семенова «Исследование и развитие метода декомпозиции для анализа больших пространственных данных» посвящена развитию метода декомпозиции для индексации, поиска и анализа больших пространственных данных. Главное внимание уделяется алгоритмам, основанным на регулярных октальных деревьях и обеспечивающим эффективное решение ряда вычислительных задач. Алгоритмы, в частности, применимы для моделирования сложных динамических пространственно-трехмерных сцен с объектами, имеющими протяженные границы. На основе вероятностного анализа выводятся оценки сложности, которые обобщают и

улучшают известные результаты и служат теоретическим обоснованием для применения алгоритмов к более широкому классу приложений.

В статье Д.Г. Федоренко и Н.А. Астраханцева «Автоматическое извлечение новых концептов предметно-специфичных терминов» описывается способ распознавания предметно-специфичных терминов, которые присутствуют в текущей базе знаний, но выражают отсутствующие в ней концепты. Разработанный метод может быть применен к неформальным базам знаний, поскольку требует только вычисления семантической близости между концептами и статистики встречаемости терминов в корпусе документов. Экспериментальная проверка показывает, что разработанный алгоритм превосходит существующие подходы, а также позволяет повысить точность разрешения лексической многозначности.

Антон Коршунов и др. в статье «Определение демографических атрибутов пользователей микроблогов» предлагают метод автоматического определения демографических атрибутов пользователей социальных сетей по текстам их сообщений и информации из профилей. Метод основан на алгоритме машинного обучения, его отличительными особенностями являются полностью автоматическое построение исходного набора данных для обучения и тестирования, а также поддержка широкого набора языков и демографических атрибутов. Экспериментальные исследования показали высокое качество результатов определения пола, возраста и семейного положения пользователя для наиболее популярных языков: английского, русского, немецкого, французского, итальянского и испанского.

Последние две статьи посвящены теоретическим основам компьютерной науки. В статье А.В. Шокурова «Нахождение корней систем алгебраических уравнений с помощью базиса Гребнера» описывается и обосновывается алгоритм нахождения решения системы алгебраических уравнений над полем k для идеалов нулевой размерности, если задан базис Гребнера идеала этой системы для определения лексикографического порядка на термах от ее переменных. Полученное решение лежит в алгебраическом замыкании основного поля. Приведен пример системы алгебраических уравнений, имеющей единственное решение в основном поле, а ее общее число решений экспоненциально относительно описания этой системы.

Наконец, в статье А.В. Воеводина и С.А. Косяченко «Оптимальное упорядочение конфликтующих объектов и задача коммивояжера» представлена постановка задачи оптимального упорядочения конфликтующих объектов и ее связь с задачей коммивояжера (Travelling Salesman Problem, TSP). Задача оптимального упорядочения конфликтующих объектов

возникает в социологии, при анализе графов в социальных сетях, при размещении рекламных заказов в сетях СМИ. Описаны используемые авторами на практике быстрые алгоритмы решения этой и связанных с ней задач. Также рассмотрена задача TSP с разреженной матрицей штрафов. Для задач TSP с ленточной и блочно-диагональной матрицами найдены необходимые и достаточные условия и построены точные алгоритмы, при которых достигается нулевое минимальное значение целевой функции задачи. Предложены эффективные алгоритмы для произвольных разреженных матриц. Приведены результаты аналитических и численных исследований сложности разработанных алгоритмов и точности решения, а также рекомендации по применению алгоритмов для решения задач подобного типа.

Академик РАН В.П. Иванников