

В 21-м томе Трудов Института системного программирования публикуются статьи сотрудников ИСП РАН, в которых описываются результаты исследований и разработок, полученные в 2011 г. Кроме того, публикуется ряд статей аспирантов ф-та ВМК МГУ, Санкт-Петербургского государственного университета, Московского государственного университета, Киевского национального университета имени Тараса Шевченко, Южно-Уральского государственного университета и Московского государственного технического университета имени Н.Э.Баумана. На английском языке представлены девять статей, которые представляют расширенные варианты докладов Седьмого ежегодного весеннего коллоквиума исследователей в области баз данных и информационных систем – организованного ИСП РАН мероприятия, ориентированного на участие молодых исследователей и разработчиков.

В статье Андрея Белеванцева, Алексея Кравца и Александра Монакова «Автоматическая генерация OpenCL-кода из гнезд циклов с помощью полиэдральной модели» предлагается способ автоматической генерации кода для стандарта OpenCL из гнезд циклов без зависимостей по данным между итерациями для программ на языках Си, Си++ и Фортран. Для генерации используется инфраструктура GRAPHITE компилятора GCC, использующая полиэдральную модель для анализа пространства итераций и пространства данных цикла.

Арутюн Аветисян, Андрей Белеванцев, Алексей Бородин и Владимир Несов в статье «Использование статического анализа для поиска уязвимостей и критических ошибок в исходном коде программ» предлагают обзор инструмента статического анализа исходного кода программ на языках Си/Си++, разработанного в ИСП РАН для поиска критических ошибок и уязвимостей. Применение межпроцедурного анализа потока данных, не гарантирующего нахождение всех заданных ситуаций, позволяет проводить автоматический анализ с долей истинных предупреждений в 40-80%.

В ИСП РАН разрабатывается инструмент статического анализа Svsace для поиска ошибок в исходном коде программ на языках Си и Си++. Цель Svsace – найти как можно больше ошибок при низком количестве ложных срабатываний и разумном использовании имеющихся ресурсов. В статье Арутюна Аветисяна и Алексея Бородина «Механизмы расширения системы статического анализа Svsace детекторами новых видов уязвимостей и критических ошибок» описывается встроенный механизм, поддерживающий включение в систему Svsace детекторов новых видов ошибок, сохраняющий ее масштабируемость.

В статье И.К. Исаева, Д.В. Сидорова, А.Ю. Герасимова и М.К. Ермакова «Avalanche: применение динамического анализа для автоматического

обнаружения ошибок в программах, использующих сетевые сокеты» рассматривается модификация и применение инструмента Avalanche для проведения динамического анализа и тестирования приложений, получающих входные данные через сокеты. Обсуждается применение модифицированной версии инструмента для анализа сетевых приложений с открытым исходным кодом, перечисляются обнаруженные во время анализа дефекты.

В настоящее время компиляторная инфраструктура LLVM не поддерживает динамический сбор профиля и перекомпиляцию, а также содержит лишь одно преобразование, использующее данные профиля. Авторами статьи «Динамическое профилирование программы для системы LLVM» А.И. Аветисяном, К. Ю. Долгоруковой и Ш. Ф. Курмангалеевым были предложены и реализованы система сбора профиля аппаратных прерываний и алгоритм, корректирующий переоценку профиля, а также несколько оптимизирующих преобразований с учетом профиля. Выполнена интеграция сбора профиля и динамического компилятора LLVM, что позволило сохранять качество программ при их переносе на другую архитектуру.

В статье А.И. Аветисяна, М.С. Акопяна и С.С. Гайсаряна «Методы точного измерения времени выполнения гнезд циклов при анализе JavaMPI-программ в среде ParJava» рассматриваются методы оценки времени выполнения модели параллельной программы на инструментальном компьютере, которые позволяют достаточно точно предсказывать время реального выполнения параллельной программы на заданном параллельном вычислительном комплексе. Модель разработана для параллельных SPMD программ с явным обменом сообщениями, написанных на языке Java с обращениями к библиотеке MPI, и включена в состав среды ParJava.

Дмитрий Мельник и Александр Монаков в статье «Поддержка команд с условным выполнением в селективном планировщике команд» предлагают метод для поддержки условного выполнения во время планирования команд, а также рассматривают преимущества данного подхода по сравнению с отдельной оптимизацией, работающей до планирования команд. Предложенный метод был реализован в селективном планировщике в компиляторе GCC. Тестирование реализации показало рост производительности на тестах SPEC FP набора SPEC CPU2000 в среднем почти на 2%.

В статье М.А. Климушеноквой и В.А. Макарова «Метод автоматического восстановления переменных из трассы исполнения программы» описывается метод восстановления локальных переменных из трассы исполнения программы. Метод использует одну из схем анализа потоков данных – достигающие определения. Рассматриваются также существующие подходы к решению задачи восстановления переменных.

Павел Довгалюк в статье «Детерминированное воспроизведение процесса выполнения программ в виртуальной машине» описывает разработку технологии, позволяющей записывать и воспроизводить сценарии выполнения программ в виртуальной машине. Данная технология позволяет выполнять детерминированную отладку приложений, а также должна в дальнейшем лечь в основу реализации различных механизмов динамического анализа программ и реверсивной отладки.

В статье П.А. Клеменкова «Оценка производительности программного обеспечения в виртуализованном окружении на основе атомарных тестов» описывается обобщенный подход к оценке требуемых программному обеспечению ресурсов при переносе его в виртуализованное окружение. Основным принцип предложенного подхода заключается в представлении сложной нагрузки в виде комбинации простых задач и замены этих простых задач на синтетические атомарные тесты.

В статье В.А. Захарова и Т.А. Новиковой «Применение алгебры подстановок для унификации программ» описывается новый алгоритм проверки логико-термальной эквивалентности программ, основанный на операции вычисления точной нижней грани в решетке конечных подстановок. На основе этого алгоритма авторам статьи удалось также решить задачу логико-термальной унификации программ, которая состоит в построении для двух заданных фрагментов программного кода такой процедуры, которая представляет собой наиболее общую специализацию этих двух фрагментов.

В.В. Липаев представил статью «Риски проектирования и производства мобильных программных продуктов», в которой вводятся основные понятия и свойства рисков комплексов программ; рассматриваются факторы и виды рисков комплексов программ и систем; обсуждается подготовка исходных данных для анализа, прогнозирования и сокращения рисков комплексов программ; описываются выделение, идентификация, анализ угроз и рисков в комплексах программ; рассматриваются методы сокращения и ликвидации опасных рисков, регистрации и утверждения допустимого интегрального риска программного продукта.

В статье И.В. Блудова «Решение проблемы NULL в запросах к реляционной базе данных на основе использования операторов реляционной алгебры A » предлагается интерпретировать логические выражения в операторах ограничения, как реляционные выражения. Автор показывает, что это позволяет по-новому подойти к решению традиционной проблемы представления в базе данных отсутствующей информации.

Целью статьи А. Эльдарханова «Обзор моделей данных объектно-ориентированных СУБД» является анализ существующих на сегодняшний день концепций формального устройства объектно-ориентированных СУБД, начиная с моделей данных и далее переходя к формальным математическим моделям (исчислением объектов, формализациям объектных языков запросов). В завершение делается заключение о наиболее актуальных проблемах моделирования ООСУБД.

В статье Д.Н. Василика «Оценка производительности протокола Snapshot Isolation» описывается реализация протокола Snapshot Isolation для распределенных СУБД, выполненная для Apache HBase, распределенного хранилища данных с открытым исходным кодом. Представлена оценка его производительности в задачах OLAP в распределенном кластере HBase.

В статье К.К. Смирнова и Г.А. Чернышева «Экспериментальное исследование параллельного исполнения SQL запросов» экспериментально изучаются два основных типа параллельного исполнения запросов – интер- и интра-операционный параллелизм и их комбинации. Эти техники применяются к дереву запроса, содержащему несколько операторов соединения, в условиях многопоточности. В экспериментах для сравнения производительности различных алгоритмов соединения варьировались количество потоков, размер буфера, характеристики тестовых данных.

Большие объемы данных, которые могут быть кластеризованы, хранятся в реляционных базах данных. Алгоритм кластеризации, реализованный на языке SQL, обеспечивает более легкий процесс кластеризации, по сравнению с использованием внешних утилит. В статье Р.М. Миниахметова «Интеграция алгоритма кластеризации Fuzzy c-Means в PostgreSQL» предложена реализация алгоритма Fuzzy c-Means, адаптированного для реляционной СУБД с открытым исходным кодом PostgreSQL.

Поисковый спам считается одной из основных угроз современным поисковым системам. Спамеры используют разнообразные методы порождения текстов, известные как текстовый спам, чтобы наполнить выдачу поисковых систем низкокачественными страницами. Методы борьбы с текстовым спамом должны основываться на большом количестве текстовых характеристик. В статье А.С. Павлова и Б.В. Доброва «Обнаружение поискового спама в Вебе на основе анализа разнообразия текстов» предлагается набор характеристик текстового разнообразия, основанных на ранговых распределениях для слов и тематик.

Извлечение информации из таблиц является важной и достаточно сложной частью информационного поиска. В статье Никита Астраханцева «Извлечение объектов и их атрибутов из таблиц текстовых документов» предлагаются

методы извлечения объектов из HTML-таблиц, решающие следующие проблемы: определение ориентации таблицы, обработка агрегирующих объектов (таких как Total) и разрозненных заголовков (подзаголовков, перерезов).

Поиск взаимосвязей между словами в тестке и статьями “Википедии”-чрезвычайно популярная задача, известная как викификация. Несмотря на её популярность, до сих пор не существует общепризнанного тестового корпуса для сравнения викификаторов. В статье С.О. Бартунова, А.А. Болдакова и Д.Ю. Турдакова «WikifyMe: создание модели сравнения для викификаторов» представлен онлайн-инструмент для совместной работы над универсальной коллекцией тестов для двух наиболее сложных задач в викификации – разрешения лексической многозначности и выделения ключевых слов.

В то время как многие исследователи пытаются построить различные онтологии с помощью Википедии, возможность получения качественных предметно-ориентированных подмножеств словаря Википедии остаётся недооценённой. В статье А.В. Коршунова, Д.Ю. Турдакова, Чингука Чонга, Минхо Ли и Чансунга Муна «Извлечение предметно-ориентированных подмножеств словаря Википедии с использованием структуры категорий» демонстрируется необходимость подобной процедуры и предлагается соответствующая методика.

В статье Мартина Давтяна «Эвристическое моделирование данных в информационных системах» описывается разработка информационных систем, которые хранят слабоструктурированные данные и используют эвристические методы для формирования гипотез относительно возможной структуры хранимых данных. Предполагается, что такие системы будут удобны в использовании, так как формализация данных будет производиться путем ответов на простые вопросы.

Том завершает статья К.С. Пана «Разработка параллельной СУБД на основе PostgreSQL» посвященная архитектуре и проектированию параллельной системы управления базами данных PargreSQL для многопроцессорных вычислительных систем с распределенной памятью. PargreSQL основана на СУБД с открытым исходным кодом PostgreSQL и использует фрагментный параллелизм.

Академик РАН В.П. Иванников