ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Арутюнян Мариам Сероповны «**Статический анализ исходного и исполняемого кода на основе поиска клонов кода»,** представленную к защите на соискание ученой степени кандидата технических наук по специальности 2.3.5 — «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

Актуальность работы

С ростом масштабов и сложности программных систем проблема дублирования кода становится все более актуальной, влияя на поддержку, безопасность и оптимизацию ПО. Стоит отметить, что поиск конов важен не сам по себе – иначе возникает вопрос о том, что делать с найденными клонами, более того, какие именно клоны надо искать (размер, семантическая замкнутость и пр.), как фильтровать несущественные варианты. Эти вопросы остро стоят в научном сообществе – за последние 20 лет было создано несколько десятков разных алгоритмов поиска клонов, и тем не менее, когда возникает потребность реализации такого алгоритма в контексте очередной практической задачи, создаётся новый алгоритм, ориентированный на особенности контекста. В целом в большом ПО могут содержаться десятки тысяч клонов, поэтому указанный выше вопрос немаловажен. На основе найденных клонов возможна дальнейшая реализация различных сервисов разработки – автоматическое уменьшение размеров кода для встроенного ПО, поиск и устранение раскопированных уязвимостей, «ручное» выделение найденных клонов в отдельные методы (в рамках автоматизированного рефакторинга в IDE) и др. Таким образом, поиск клонов с их последующей автоматизированной обработкой является актуальной задачей.

Краткое описание работы

Диссертация состоит из введения, пяти глав и заключения.

Во введении изложена цель работы, обоснована ее актуальность, даны определения ключевых понятий и сформулированы результаты, выносимые на защиту.

В первой главе представлен обзор существующих работ, относящихся к теме диссертации. Подробно рассматриваются и сравниваются современные методы поиска клонов в исходном и исполняемом коде, инструменты сравнения программ и идентификации библиотек, а также существующие исследования по обнаружению циклической проверки избыточности (ЦПИ).

Во второй главе представлен метод поиска клонов, основанный на анализе графов зависимостей программы.

В третьей главе рассмотрен метод оптимизации программ путем замены неоптимальных реализаций ЦПИ на более эффективные.

В четвертой главе предложен двухэтапный метод выявления изменений между версиями программ, сочетающий метрический подход и разработанный метод поиска клонов кода.

В пятой главе описаны методы идентификации статически связанных библиотек и поиска уязвимостей в программном коде.

В заключении приводятся результаты диссертационной работы.

Научная новизна и основные результаты

Автором предложены новые методы, перечисленные ниже.

- Метод поиска клонов в исходном и исполняемом коде на основе графов зависимостей программы, обеспечивающий высокую точность и масштабируемость.
- Метод оптимизации кода, основанный на замене обнаруженных клонов неэффективных реализаций ЦПИ на более оптимальные версии.
- Двухэтапный метод выявления изменений между версиями программ, включающий метрический подход и анализ графов зависимостей.
- Методы идентификации статически связанных библиотек и поиска известных уязвимостей, позволяющие анализировать программные зависимости и улучшать безопасность кода.

Эти методы обладают определённой теоретической новизной и существенной практической значимостью, позволяя улучшить существующие инструменты анализа программного кода и автоматизации разработки ПО.

Достоверность и обоснованность результатов

Достоверность результатов подтверждается экспериментальными исследованиями, проведенными на реальных проектах (OpenSSL, Coreutils, Binutils и др.), а также сравнением с существующими подходами.

Основные результаты докладывались и обсуждались на 9 международных научных конференциях. По теме диссертационной работы автором опубликовано 12 статьей, включая четыре статьи в рецензируемых журналах, рекомендованных ВАК РФ. Четыре работы проиндексированы в базе данных Scopus, две статьи опубликованы в журналах Q1 (версия SJR). Кроме того, получено два свидетельства о регистрации программ для ЭВМ.

Теоретическая и практическая значимость

Теоретическая значимость диссертации заключается в развитии методов анализа программного кода, включая использование графов зависимостей и техник машинного обучения для поиска клонов. Практическая ценность подтверждается внедрением разработанных методов в платформы анализа кода GenesISP и BinSide, а также возможностью их применения в компиляторе GCC.

Замечания

1. Наличие к настоящему моменту большого числа методов и алгоритмов поиска клонов наводит на мысль о контекстно-зависимой природе данной задачи и принципиальной сложности универсальных решений. Логика автора про то, что в имеющихся подходах имеются сложности с поиском нечётких клонов (тип 3) и семантических клонов (тип 4), которые она взяла и решила прямолинейна. На мой взгляд, следовало делать акцент на те сервисы, в рамках которых используется поиск клонов в работе, поскольку именно этот контекст является уникальным и формулирует особые требования к задаче. Здесь, на мой взгляд, находится новизна.

2. В подразделе 4.1.2. указано, что веса w_i были установлены с помощью методов машинного обучения, однако не раскрываются детали этого процесса. Необходимо пояснить, каким образом применялись методы машинного обучения, как происходило обучение и оценка полученных результатов.

3. В работе лишь упомянута группа в JetBrains, занимающаяся поисками клонов, но не описан сам подход, реализованный в продукте Intellij IDEA как

составная часть рефакторинга (выделения в метод). А между тем данный подход имеет ряд свойств, значимых для данной диссертационной работы на работу с большой кодовой базой, мультиязыковость и др.

Однако указанные замечания не снижают значимости полученных результатов и не влияют на общую положительную оценку диссертационного исследования.

Заключение

Диссертация Арутюнян Мариам Сероповны является завершенной научно-исследовательской работой, содержащей теоретические и практические результаты, имеющие важное значение для области анализа программного кода. Работа соответствует требованиям ВАК, предъявляемым к кандидатским диссертациям, а автор заслуживает присуждения ученой степени кандидата технических наук.

Профессор кафедры системного программирования Санкт-Петербургского государственного университета, доктор технических наук /Кознов Дмитрий Владимирович/

27.03.2015