

ОТЗЫВ

официального оппонента на диссертационную работу Луцива Дмитрия Вадимовича на тему «Поиск неточных повторов в документации программного обеспечения», представленную на соискание учёной степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Диссертация посвящена проблеме разработки и сопровождения документации программного обеспечения (ПО), которая не теряет актуальности на протяжении последних 50-ти лет. Существует большое разнообразие подходов к решению этой проблемы, однако их применимость на практике зависит от большого количества параметров ситуации – типа проекта и программного обеспечения, масштаба разработки, вида организации, используемой в проекте методологии разработки, особенностей организационной культуры и т.д. При этом даже сам феномен остаётся плохо определённым: существуют многочисленные различные классификации видов документации ПО, и точный состав пакета документации программного продукта каждый раз определяется, исходя из различных факторов и обстоятельств. С другой стороны, наблюдается тенденция усложнения средств разработки ПО – все более многофункциональные IDEs, все более сложные средства поддержки процесса разработки и отдельных видов деятельности (тестирования, управления версиями и пр.). В связи с этим в сферу разработки документации уже давно проникают инструменты, напоминающие средства программирования, в частности, средства повторного использования документации. При этом, поскольку на практике приходится иметь дело с уже существующей документацией, в контексте повторного использования актуальной является задача поиска существующих повторов. Именно этой задаче и посвящена предлагаемая диссертационная работа.

За последние 10 лет опубликован ряд исследований, посвящённых поиску повторов в документации ПО. Тем не менее, следует отметить, что эти исследования носят преимущественно характер case-study, не содержат определений, моделей и алгоритмов, и, таким образом, не продвинулись дальше изучения повторов в отдельных конкретных документах. Более того, отмечается важность неточных повторов, однако их поиск продолжал оставаться нерешённой задачей.

Основные результаты диссертации заключаются в следующем.

1. Предложена модель неточных повторов программной документации. Модель основана на концепции архетипа Пауля Бассета. Приводится два определения неточных повторов – первое ориентировано на компоновку неточных повторов из точных, а также имеет фиксированную константу сходства, второе определение понимает под повторами произвольные «близкие» текстовые фрагменты, для которых отношение фрагмента к некоторому фиксированному архетипу (с разрывами) не превосходит фиксированного числа в диапазоне от 0 до 1 (мера близости данной группы повторов). Модель является новой, так как до сих пор в литературе были представлены лишь case studies на эту тему, и формальные определения неточных повторов и сопутствующих объектов в них отсутствовали. Известна единственная работа (М. Носал и Я. Порубан, 2016 год), которая содержит явное определение неточного повтора в документации ПО, но это определение является неформальным.
2. Предложен алгоритм поиска неточных повторов на основе компоновки точных с помощью средств поиска клонов в программном обеспечении. Алгоритм построен на основе поиска точных повторов, осуществляемого средствами поиска клонов ПО. Новизна алгоритма заключается в том, что он ищет неточные повторы в документации ПО, что не могли делать представленные до сих пор в литературе средства. Также важной деталью является тот факт, что алгоритм формализован, в работе представлено доказательство его корректности относительно представленной выше модели неточных повторов. В тоже время существующие средства поиска повторов в документации ПО являются неформальными. Наконец, важным достоинством алгоритма является то, что он реализован в рамках программного пакета Duplicate Finder, разработанного диссертантом.
3. Предложена интерактивная методика поиска неточных повторов, позволяющая учитывать заданную экспертом семантику повторов. В рамках методики разработан алгоритм поиска по образцу, доказана полнота этого алгоритма. Интересным является применение метафоры карты нагрева (Heat Map) для навигации пользователя по документу и выбора фрагмента-образца для дальнейшего поиска. Карта строится на основе повтора отдельных слов документа – учитывается частота вхождения слова в группы точных повторов, найденные с помощью поиска точных повторов средствами поиска клонов в ПО. Автор доказал, что выдача алгоритма содержит (пересекает с хорошей долей покрытия) все неточные повторы выбранного образца, содержащиеся в

документе. Найденные неточные повторы можно редактировать (увеличивать/уменьшать размеры), а также исключать из списка кандидатов нерелевантные текстовые фрагменты.

4. Создан метод улучшения документации на основе неточных повторов, выполняющий. Важно отметить, что метод включает в себя рефакторинг документации в формате DocBook, то есть поддерживает автоматизированное выделение описаний повторно используемых фрагментов и замену их вхождений на вызовы с параметрами созданных описаний. При этом производится балансировка разметки документа и соответствующих текстовых фрагментов (т.е. вставка открывающих закрывающих/тегов). Предложенный метод отличается от аналогов тем, что утверждает полезность повторов в документации, в то время как большинство имеющихся исследований считают повторы избыточностью и плагиатом. С другой стороны, имеются подходы, которые ориентируются на повторное использование документации, однако они не исследуют «природу» повторов. Автор диссертационного исследования рассматривает так называемую ссылочную документацию (reference), введенную в 2011 году Давидом Парнасом и включающую такие виды документации, как спецификации программных интерфейсов, руководства пользователей, руководства по языкам программирования и т.д., т.е. такие документы содержат описание однотипных объектов и предназначены для использования в режиме справочника. Автор утверждает, что повторы в такой документации крайне желательны и способствуют унификации документации. Предложенный метод нацелен на модификацию такой документации с целью создания максимального числа повторов для исправления ошибок и неточностей, а также для её унификации. Представлена апробация метода на ряде документов известного открытого программного обеспечения.

Следует отметить, что представленные результаты обладают научной новизной и в должной мере обоснованы – как с помощью формальных доказательств, так и посредством экспериментов. Результаты исследования прошли апробацию на конференциях и семинарах, как в России, так и за рубежом. Все результаты работы опубликованы в 8 печатных работах, из них 3 статьи изданы в журналах из “Перечня российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание учёных степеней доктора и кандидата наук”, 3 статьи опубликованы в изданиях из баз цитирования SCOPUS и Web of Science. Для статей, опубликованных в соавторстве, в диссертации и в

автореферате описан личный вклад автора. Автореферат полностью отражает содержание диссертации и, как и диссертация, исчерпывающе описывает вклад автора в разработку представленных им результатов.

Значимость для практики полученных результатов и выводов диссертанта заключается в создании средств для поиска нечётких в документации ПО. Причём эти средства реализованы в рамках программного инструмента, который имеет хорошо проработанный пользовательский интерфейс, основанный на оригинальных метафорах визуализации для анализа многомерных данных о повторах в документации. Программный инструмент реализован на языке Питон, его исходный код выложен в Интернет, в свободный доступ. Инструмент, фактически, готов для использования в индустрии. Для практики также важен предложенный автором метод улучшения документации на основе неточных повторов.

Диссертация состоит из введения, пяти глав, заключения и приложения. Полный объем диссертации составляет 122 страницы, включая 16 рисунков и 5 таблиц. Список литературы содержит 153 наименования.

Замечания по содержанию и оформлению диссертационной работы.

1. Работу, вне всякого сомнения, украсили бы теоретические оценки сложности разработанных алгоритмов.

2. Терминология последней главы, посвященной описанию инструмента, могла бы быть более стандартной, то есть следовать существующим ГОСТам и известным международным рекомендациям по разработке ПО.

3. Предложенный автором метод улучшения документации на основе поиска неточных повторов ориентирован лишь на существующую документацию, однако было бы интересно рассмотреть также ситуацию создания большого пакета ссылочной документации «с нуля». Этот важный вопрос оказался за рамками представленного диссертационного исследования.

Отмеченные недостатки, однако, не носят принципиального характера и не ставят под сомнение основные результаты, полученные в диссертационной работе.

Диссертация Д.В. Луцива соответствует паспорту специальности 05.13.11: методы, модели и алгоритмы проектирования и анализа программ и программных систем, их эквивалентных преобразований, верификации и тестирования. Работа полностью удовлетворяет требованиям Положения "О порядке присуждения учёных степеней», предъявляемым к кандидатским диссертациям, а её автор, Луцив Дмитрий Вадимович, заслуживает присуждения учёной степени кандидата физико-математических наук по

специальности 05.13.11 — математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Официальный оппонент:

Профессор кафедры вычислительной техники
ФГБОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)»

Доктор технических наук,
Профессор

Водяхо Александр Иванович

«21» марта 2018 г.

Сведения о составителе отзыва:

ФИО: Водяхо Александр Иванович

Ученая степень: доктор технических наук

Ученое звание: профессор

Место работы: Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)»

Должность: профессор кафедры вычислительной техники

Почтовый адрес: 197376, Россия, Санкт-Петербург, ул. Профессора Попова, дом 5.

Телефон: (812) 234-25-03

Адрес электронной почты: aivodyaho@mail.ru

Подпись А.И. Водяхо заверяю:

Начальник отдела диссертационных советов



Русяева Т.Л.

2018 г.