

## ОТЗЫВ

официального оппонента на диссертационную работу  
**Луцива Дмитрия Вадимовича**  
**«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»

Разработка промышленной документации программного обеспечения неизменно привлекает внимание исследователей в области программной инженерии. Не утихают споры о том, какую именно документацию следует создавать при разработке программных систем и сколь подробной она должна быть. При этом продолжают активно создаваться и совершенствоваться языки разработки документации, различные методы и подходы к её сопровождению. Одной из интересных тем исследования в этой области является повторное использование документации. В свете этого направления возникает задача поиска повторов и, в частности, неточных повторов, в уже существующей документации. Данной задаче посвящён ряд исследований, однако методов поиска неточных повторов до сих пор не было предложено. Данная диссертация посвящена решению этой задачи.

В ходе диссертационного исследования были получены следующие основные результаты.

1. Формальная модель неточных повторов в программной документации, алгоритм поиска неточных повторов в документации ПО на основе компоновки точных повторов, найденных с помощью метода поиска точных клонов ПО; доказана корректность алгоритма.
2. Методика интерактивного поиска неточных повторов, позволяющая учитывать заданную экспертом семантику повторов, алгоритм поиска по образцу; доказана полнота данного алгоритма.
3. Метод улучшения документации на основе неточных повторов, включая средства автоматизированного рефакторинга документации в формате DocBook.

Данные результаты обладают научной новизной, поскольку решают нерешённую до сих пор задачу поиска неточных повторов. Дело в том, что существующие работы использовали имеющиеся средства поиска повторов (главным образом, средства поиска программных клонов), которые хорошо справлялись с задачей поиска точных повторов, но требовали серьёзной и, во многих случаях, неочевидной доработки для поиска неточных повторов. Именно такую доработку средств поиска программных клонов и выполнил Д.В.Луцив. Кроме того, он использовал известную парадигму поиска по образцу и предложил простую методику для семантического поиска неточных повторов, реализуя ее на основе интерактивности (взаимодействия с пользователем). Несмотря на то, что существуют серьёзные средства семантического поиска, такой облегчённый подход позволил быстро получить значимые для данной области результаты.

Требуется отметить, что представленные результаты тщательно обоснованы – автор построил формальную модель нечётких повторов и относительно неё формально доказал корректность и полноту предложенных алгоритмов. Также было проведено значительное количество экспериментов, показывающих, что созданные алгоритмы и методика могут использоваться на практике. Наконец, автор выполнил программную реализацию предложенных им результатов, создав интересные метафоры визуализации многомерной информации о повторах в документации. (К сожалению, эти метафоры не были включены в результаты работы.)

Материал диссертации достаточно полно отражён в опубликованных автором работах. По теме диссертации опубликовано 8 печатных работ, из них 3 статьи изданы в российских рецензируемых научных журналах, 3 статьи – в изданиях, входящих в реферативные базы данных Web of Science и Scopus.

Текст диссертация состоит из введения, пяти глав, заключения и приложения. Его объём составляет 122 страницы, включая 16 рисунков и 5 таблиц; список литературы содержит 153 наименования.

Первая глава посвящена обзору области исследования. Во-первых, кратко рассмотрена проблема разработки документации ПО, во-вторых, описаны ряд существующих классификаций документации ПО, в-третьих, описаны имеющиеся на сегодняшний день исследования в области поиска повторов в документации ПО, в-четвертых, сделан обзор markup-языков разработки документации, в-пятых, представлены методы, модели и технологии, используемые в диссертации. Наконец, в этой главе автор приводит результаты небольшого собственного исследования о размере документации ПО и представляет набор документов, которые используются в дальнейшем для экспериментов.

Во второй главе представлены формальные определения неточных повторов, алгоритм компоновки неточных повторов и экспериментальные данные о его



эффективности. Также доказана корректность предложенного алгоритма относительно модели нечётких повторов. Представлены результаты экспериментов, показывающие, что алгоритм имеет приемлемое быстродействие.

Третья глава описывает методику интерактивного поиска неточных повторов, включающую алгоритм поиска по образцу. Данная методика, через взаимодействие с пользователем, реализует семантический поиск повторов: пользователь управляет поиском, выбирая образцы и корректируя выдачи алгоритма. Также сформулирован критерий полноты и доказана полнота предложенного алгоритма. Представлены результаты экспериментов, показывающие, что алгоритм поиска по образцу имеет приемлемое быстродействие.

В четвертой главе представлен метод улучшения документации на основе неточных повторов. Данный метод обеспечивает на поиск и анализ существующих повторов, и позволяет исправлять ошибки и выполнять унификацию анализируемой документации.

В пятой главе представлен инструмент Duplicate Finder, реализующий представленные алгоритмы и методику. Инструмент поддерживает два режима: автоматический и интерактивный. Автоматический режим позволяет сделать быструю оценку картины повторов в документации, используя алгоритм компоновки. Интерактивный режим позволяет выполнить более продвинутый поиск неточных повторов, учитывая их семантику. Инструмент поддерживает два формата: DocBook и «плоский» текст. Инструмент разработан на языке Питон, его исходный коды опубликованы в Интернете.

В качестве **замечаний** можно выделить следующее.

1. На мой взгляд, описание документов, которые используются в экспериментах диссертационного исследования, содержится в неподобающем месте – в первой главе (лит. обзор). При этом сами эксперименты разбросаны по всей диссертации. А в автореферате, при описании последней главы, сказано, что она содержит описание экспериментов, но в тексте диссертации их там нет. Видимо автор справедливо хотел поместить все эксперименты в одно место, но по какой-то причине не сделал этого.
2. В работе отсутствуют теоретические оценки сложности предложенных алгоритмов, хотя представляется, что их не сложно было выполнить.

Указанные недостатки не умаляют значимости данной работы. Работа имеет несомненную научную и практическую ценность. Автореферат в полной мере отражает содержание диссертации.

Таким образом, можно заключить, что диссертационная работа «Поиск неточных повторов в документации программного обеспечения» соответствует критериям,

установленным Положением «О присуждении 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.

Кандидат технических наук, доцент,  
доцент высшей школы программной инженерии  
Федерального государственного автономного  
образовательного учреждения высшего образования  
«Санкт-Петербургский политехнический университет  
имени Петра Великого»

Дробинцев Павел Дмитриевич

*Дробинцев П.Д.* 28.03.2012

