

ОТЗЫВ

научного руководителя на диссертацию Дмитрия Вадимовича Луцива **«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 документации программного обеспечения (ПО) была и остаётся нерешённой научно-практической проблемой. О наличии этой проблемы исследователи пишут с 70-х годов (например, Ф.Брукс-младший в начале 70-х) до настоящего времени (например, Дэвид Парнас в 2011 году). Трудность заключается в отсутствии общепринятых взглядов на эту проблему, в виду большого разнообразия видов ПО, а также из-за изменчивости ПО (то есть большого объёма модификаций). Последнее обстоятельство требует специальных методов и технологий для анализа, прослеживаемости и поддержания целостности документации, которая должна модифицироваться вслед за изменённым ПО.

В работе Дмитрия Вадимовича выдвигается тезис о том, что поиск и анализ повторов в документации ПО очень полезен для исправления ошибок, а также может внести существенный вклад в стандартизацию документации. То есть, в отличие от других исследователей, которые пишут о повторах как о плагиате и ненужной избыточности, автор считает повторы положительной тенденцией. В работе предложен метод, который на основе анализа существующей документации позволяет выполнять стандартизацию документации, максимально увеличивая количество повторов при представлении одной и той же информации. При этом автор делает акцент на справочной (reference) документации, введённой Д.Парнасом (речь идёт о руководствах пользователя, API-документации и пр.). Наконец, в своём исследовании Дмитрий Вадимович рассматривает неточные повторы, что, фактически, сделано впервые. Необходимость находить и анализировать неточные повторы при сопровождении документации ПО отмечалась многими авторами, но реальные шаги в этом направлении были крайне

незначительны: были предложены лишь неформальные определения неточных повторов, а также выполнен поиск таких повторов на очень специальных частных случаях документации ПО.

В своей диссертационной работе Дмитрий Вадимович предлагает формальную модель неточных повторов, основанную на работах Пола Бассета, а также ряд алгоритмов для поиска таких повторов в документации ПО, для которых он формально доказывает корректность и полноту. Автор предлагает набор оптимизаций для этих алгоритмов и ставит ряд экспериментов, которые показывают, что эти алгоритмы применимы на практике. В работе предложена методика поиска неточных повторов на основе тепловых карт (Heat Map) и интерактивного поиска по образцу, что позволяет до некоторой степени решить вопрос семантической осмысленности находимых повторов (решение строится на взаимодействии с пользователем – то есть именно последний решает, что осмысленно, а что нет, и управляет поиском). В работе предлагается автоматизированный рефакторинг Docbook-документации на основе найденных повторов с частичным решением проблемы балансировки markup-разметки. Дмитрий Вадимович предложил также интересный метод улучшения и стандартизации документации на основе неточных повторов. Наконец, в работе представлен программный инструмент, реализующий предложенные автором идеи. Инструмент содержит ряд оригинальных идей по визуализации информации о неточных повторах и выполнен на высоком профессиональном уровне. Его реализация выложена в открытый доступ.

Необходимо отметить, что работа является многокомпонентой. Она принадлежит к области программной инженерии, но включает в себя элементы информационного поиска и теоретической информатики. Иными словами, работа содержит формальные определения, теоремы и доказательства, а также алгоритмы и эксперименты, инженерный метод, и, наконец, интересный программный инструмент. Всё это разнообразие результатов Дмитрию Вадимовичу удалось увязать в целостную работу, предварив исследование интересным анализом проблем, методов и средств разработки документации ПО. Таким образом, хочется отметить высокую эрудированность автора в программной инженерии и смежных областях, а также солидный практический опыт в области индустриального программирования.

Следует также отметить, что данная работа вписана в контекст международного исследовательского сообщества по программной

инженерии. Несмотря на то, что результаты не были пока ещё опубликованы в ведущих конференциях ICSE, ASE, ICSR, они были опубликованы в Springer (2 статьи). Также автор докладывал свои результаты на международных конференциях PSI 2015, SYRCoSE 2017, APSSE 2017.

Работа имеет как теоретическое, так и практическое значение. Теоретическое значение заключается в построении формальной модели неточных повторов, в формальном изложении разработанных алгоритмов, а также в наличии доказательств их основных свойств. Практическое значение заключается в создании методики, метода и соответствующего программного инструмента, которым можно пользоваться в реальных проектах. Предложенный метод вызвал оживлённую дискуссию на конференции APSSE 2017, которая показала актуальность данного исследования и реальный практический интерес к его результатам со стороны индустрии. Я считаю, что у представленного исследования имеются реальные перспективы внедрения при разработке документации крупных программно-аппаратных комплексов.

Таким образом, диссертацию можно считать законченной научно-исследовательской работой, её тематика и полученные научные результаты полностью соответствуют паспорту специальности 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей». Работа соответствует требованиям Положения о порядке присуждения учёных степеней, а её автор, Дмитрий Вадимович Луцив, заслуживает присуждения ему учёной степени кандидата физико-математических наук по специальности 05.13.11.

Доктор технических наук,
профессор кафедры системного
программирования Санкт-петербургского
государственного университета

Д.В.Кознов

06.02.2018

