

Отзыв научного руководителя на магистерскую выпускную работу

обучающейся Коноваловой Ирины Михайловны

Поиск и кластеризация нечётких повторов в документации программного обеспечения

На всём протяжении истории программной инженерии, по мнению исследователей и представителей индустрии, документация в софтверных проектах традиционно является одним из недооценённых и проблемных активов вследствие дефицита внимания и культуры работы с ней. Повторяющийся текст в документации часто является причиной ошибок и общего снижения качества документации, так как повторы не отслеживаются и не контролируются должным образом.

Ранее коллективом исследователей с отделения информатики Математико-механического факультета СПбГУ были опубликованы работы по тематике управления повторами в документации, причём данные, использованные для апробации предложенных в этих работах подходов и алгоритмов, обычно собирались и размечались экспертами (людьми) на предмет повторов в полуавтоматическом режиме. Это позволяло гарантировать осмысленность выделенных повторов и групп, но, из-за привлечения экспертов, достаточно дорогой ценой. В своей работе Ирина Михайловна воспользовалась тем, что справочная документация как правило уже разбита на осмысленные фрагменты и тем, что в датасетах предыдущих работ уже накопилось достаточно информации о том, как такое разбиение выполняет человек. В итоге ею были проанализированы различные виды векторных представлений, алгоритмы кластеризации и параметры этих алгоритмов с целью нахождения такого их сочетания, которое давало бы результат наиболее похожий на результат экспертов и обеспечивало бы высокую производительность. Результат работы Ирины Михайловны позволит упростить дальнейшие исследования по тематике документации программного обеспечения.

Для решения поставленной задачи автором был реализован ряд фильтров для импорта размеченной и неразмеченной документации, а также выполнены эксперименты. При планировании и выполнении экспериментов Ирина Михайловна продемонстрировала высокий уровень культуры и самостоятельности работы, показав себя квалифицированным исследователем и эрудированным специалистом по анализу данных.

Текст работы написан понятным языком, хотя структура его на первый взгляд может показаться неочевидной: при наличии постановки задачи основные разделы текста работы структурированы не по ней, а по вопросам исследования. Но это можно объяснить тем, что работа носит не инженерный, а исследовательский характер, и главный смысл несут ответы на вопросы, а результаты лишь фиксируют достигнутое. Ряд мелких огрехов в оформлении текста не мешают его восприятию и не снижают общего впечатления от него.

Считаю, что автору имеет смысл продолжить образование и исследования в аспирантуре.

Рекомендую присвоить Ирине Михайловне квалификацию магистра и оценить работу «отлично».

Научный руководитель:

к.ф.-м.н., старший преподаватель кафедры системного программирования СПбГУ

Луцев Дмитрий Вадимович



6 июня 2020 г.