

Отзыв научного руководителя на магистерскую выпускную работу

обучающегося Глазырина Антона Георгиевича

Универсальный механизм первичного поиска повторов в тексте для пакета Duplicate Finder

На всём протяжении истории программной инженерии, по мнению как исследователей, так и представителей индустрии, документация в программных 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.

В последнее десятилетие сотрудниками и обучающимися Кафедры системного программирования разрабатывается инструмент Duplicate Finder, предназначенный для автоматического и интерактивного поиска и работы с неточными повторами в документации различных форматов. Любопытным фактом является то, что в основе инструмента на протяжении всего его существования лежал другой инструмент — предназначенный для поиска повторов в программном коде Clone Miner, задуманный нашим коллегой профессором Хамидом Абдулом Баситом из университета Лахора (Пакистан), и реализованный им и его студентами.

Clone Miner неплохо справлялся со своими задачами, но, не будучи предназначенным для текстов на естественных языках, в ряде случаев требовал дополнительной предобработки текста документации и постобработки результатов поиска. Кроме того, исходный код инструмента был закрыт, вследствие чего авторы Duplicate Finder не могли сопровождать его.

Перед Антоном Георгиевичем была поставлена задача реализовать инструмент, который, с одной стороны, мог бы заменить Clone Miner, а с другой — использовать традиционные именно для работы с текстом на естественном языке техники. Как и остальной Duplicate Finder, этот инструмент в его составе реализован на Python, что облегчает его поддержку и делает его интеграцию в Duplicate Finder более изящной и легковесной.

Антон Георгиевич справился с поставленной задачей: практическая часть была реализована, протестирована и оформлена в виде запроса на слияние в репозиторий исходного кода Duplicate Finder, а описательная — в виде хорошо структурированного, аккуратно оформленного и понятно написанного текста.

Впечатление от работы Антона Георгиевича несколько омрачает тот факт, что она не была завершена годом раньше: для этого были все предпосылки, но автор решил не торопить события. Впрочем, это положительно сказалось на качестве результата. Также, а мой взгляд, Антон Георгиевич напрасно пренебрёг возможностью представить свой результат на одной из подходящих как по тематике, так и по уровню работ, конференций, например, СПИСОК-2023 в СПбГУ или «Современные технологии в теории и практике программирования» в Политехническом университете.

Указанные недостатки не являются, тем не менее, недостатками работы, как таковой. Поэтому считаю, что выполненная работа позволяет присудить Антону Георгиевичу квалификацию магистра по направлению «Программная инженерия», и рекомендую комиссии оценить работу «отлично» (А по системе оценивания ECTS).

Научный руководитель:

к.ф.-м.н., доцент кафедры системного программирования СПбГУ

Луцив Дмитрий Вадимович



1 июня 2023 г.