

Отзыв

научного руководителя на работу студентки Ирины Коноваловой Коноваловой «Разработка эталонной коллекции для задачи поиска нечётких дубликатов в документации программного обеспечения»

Работа Ирины посвящена изучению неточных повторов в документации программного обеспечения. В работе было сделано следующее.

1. Для API-документации LibLDAP Manual и Python Requests были созданы эталонные коллекции неточных повторов (76 и 31 групп соответственно).
2. Спроектирована и реализована система для работы с эталонными коллекциями неточных повторов на языке Python в контексте инструментария Duplicate Finder, включая разработку метрик оценки качества точность (Precision), полнота (Recall), средняя точность (Average Precision), средняя полнота (Average Recall).
3. Выполнена апробация созданного подхода и разработанной системы на алгоритме компоновки пакета Duplicate Finder. Было выявлено, что почти все повторы, содержащиеся в выдаче алгоритма, являются релевантными, но алгоритм компоновки часто некорректно выделяет границы найденных повторов. Кроме этого алгоритм «склеивает» фрагменты текста, находящиеся рядом, но описывающие разные программные сущности (software features), в результате чего теряется большое количество релевантных повторов.

Работа интересная и многосторонняя – Ирина разобралась с несколькими областями (проект Duplicate Finder, программная документация ряда открытых проектов, информационный поиск и бенчмарки) и сумела довести работу до конца, получив при этом интересные результаты. Рекомендую работу к публикации, а самой Ирине рекомендую оценку «отлично».

Доктор технических наук,
профессор кафедры системного программирования Кознов Д.В.