

ОТЗЫВ

научного руководителя на выпускную работу
бакалавра кафедры Технологии Программирования
Санкт-Петербургского государственного университета
Зайцева Андрея Алексеевича

Кластеризация с поиском дубликатов на примере патентов

Известно, что патенты содержат огромный объем информации, которая часто не публикуется где-либо еще. Также из-за большого объема патентной базы и как следствие сложности точного анализа нередко случаются схожести описания патентов. В связи с этим решение задачи поиска дубликатов и кластеризации патентов является весьма актуальной.

Задача, решению которой посвящена работа Зайцева А.А., состоит из технической и исследовательской частей. В исследовательской части Андреем были приведены описания алгоритмов кластеризации (Canopy Clustering и Mini-Batch K-Means) и, также, критерии их оценок. В работе вводится новая мера сходства и критерий согласно которому документ считается дубликатом. Следует отметить математическую строгость и грамотность оформления материала. К технической части относится формирование и индексирование большой коллекции патентов (более 400,000) их анализ на предмет подозрения на дубликат. Андрей продемонстрировал умение работать с научной и справочной литературой, т.к. для проведения данного исследования была подробно изучена и критически проанализирована большая база источников преимущественно на иностранном языке.

В целом, можно отметить, что Андрей показал себя серьезным исследователем, способным решать большие реальные задачи. Полагаю возможным оценить его квалификационную работу «Кластеризация с поиском дубликатов на примере патентов» оценкой **«отлично»**.

Ст. преподаватель

А.В. Уланов