

РЕЦЕНЗИЯ

на выпускную квалификационную работу Огурцовой Анастасии Сергеевны **Автоматическое составление обзора важнейших событий на основе анализа русскоязычных новостных лент**

В выпускной квалификационной работе успешно решена задача разработки программы, составляющей обзор событий из имеющейся коллекции новостных документов.

Во введении автор отмечает, что количество информации, с которым человек сталкивается ежедневно, неуклонно растёт и обработать её вручную становится практически невозможно, да и человеческих ресурсов потребовалось бы слишком много. Человек уже не в состоянии проанализировать множество имеющихся новостных потоков. Разработка методов для автоматической обработки информации позволяет сократить объём информационного потока до разумных размеров. С применением автоматического анализа, человеку не требуется просматривать огромное количество новостных материалов для того, чтобы оставаться в курсе последних событий. Таким образом, автор делает вывод о том, что задача автоматической обработки новостных статей является востребованной и актуальной.

В первой главе автор говорит о том, что новостные порталы размещают за очень короткое время множество статей, посвященных одному и тому же значимому событию. Поэтому, в первую очередь, требуется объединить новостные публикации, соответствующие одному событию, в группы. Такая задача называется задачей кластеризации. Автор рассматривает различные методы кластеризации, которые включают иерархические, статистические, вероятностные, графовые методы, а также метод главных компонент.

Вторая глава посвящена составлению обзора событий. Для извлечения наиболее значимых предложений, можно представить исходный текст в виде графа. Вершины графа - предложения исходного текста. Наиболее важными считаются те предложения, которые содержат информацию из нескольких других предложений. При этом вес рёбер отражает связь соответствующей пары предложений. Затем, основываясь на количестве входящих ребер и их рейтинге, вычисляется рейтинг каждой вершины. После этого, полученные веса вершин применяются для извлечения сведений из текста. В этой же главе рассматриваются алгоритмы, основанные на деревьях принятия решений.

Третья глава содержит практические результаты работы и данные, полученные в ходе экспериментов. Для составления обзора событий автором был изучен и применён алгоритм TextRank. Однако, данный метод предполагает знание количества требуемых предложений, что не всегда возможно. В связи с этим, задача составления обзора событий была сведена к задаче бинарной классификации. Для её решения, используя алгоритм CART, был обучен классификатор, принимающий решения по включению рассматриваемого предложения в обзор. Помимо этого был также обучен классификатор, основанный на алгоритме Random Forest. Однако, он показал не такие хорошие результаты. В итоге автор делает вывод о том, что, наилучшей комбинацией методов для решения поставленной задачи является алгоритм k-means++, использованный совместно с методом главных компонент, и классификатор, использующий метод CART.

В результате выполнения данной работы была разработана программа, которая составляет обзор событий для имеющейся коллекции русскоязычных новостных

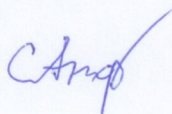
документов. Кроме этого были размечены тестовая выборка для оценки качества кластеризации и обучающая выборка предложений для классификации, а также рассмотрены популярные алгоритмы машинного обучения.

В целом выпускная квалификационная работа Огурцовой А.С. является комплексной законченной работой. Автор свободно владеет терминологией и математическим аппаратом, показывает свои навыки в разработке программных продуктов. Список использованной литературы отражает хороший уровень проработки известных результатов в области исследований. В работе имеются необходимые ссылки. Структура работы соответствует логике исследования.

Таким образом, выпускная квалификационная работа Огурцовой Анастасии Сергеевны.

“Автоматическое составление обзора важнейших событий на основе анализа русскоязычных новостных лент” заслуживает оценки **отлично**.

Кандидат технических наук



Анфиногенов Сергей Олегович

26 мая 2017