

РЕЦЕНЗИЯ

на выпускную квалификационную работу Логачева Михаила Максимовича
Детектирование общественно значимых новостей в потоке сообщений

В выпускной квалификационной работе успешно решена задача выделения актуальных событий из потока сообщений.

Во введении автор отмечает, что большинство новостей люди узнают, прежде всего из новостных лент социальных сетей. Часто это сообщения от официальных аккаунтов известных новостных агентств и изданий. Но в социальных сетях также публикуется информация от очевидцев событий, людей, которые оказались рядом с местом события или же являются непосредственными их участниками. С помощью данной информации новостные агентства и экстренные службы могут быстрее узнавать о событиях, и реагировать на происходящее. Таким образом задача выделения аномальных и общественно значимых событий из потока сообщений в социальных сетях, а также их последующий анализ является востребованной и актуальной.

Первая глава посвящена сбору сообщений для анализа. Автор рассказывает об особенностях построения системы сбора сообщений и отмечает, что для сбора достаточного количества данных требовалось обеспечить бесперебойную работу программного продукта и базы данных для хранения полученной информации, поэтому для размещения необходимых систем и программ было решено использовать облачный сервис.

Во второй главе автор говорит о том, что перед проведением исследования и сравнения различных методов кластеризации необходимо привести собранные сообщения к векторному представлению, чтобы уже по их векторному виду проводить кластеризацию и дальнейший анализ. Рассматривается два варианта преобразования сообщений к векторному виду. Как векторное преобразование для всех слов в сообщениях, так и построение векторного представления лишь по существительным и именованным сущностям так как была выдвинута гипотеза о том, что информация обо всех словах может быть избыточной.

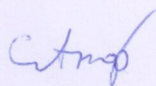
В третьей главе автор рассматривает четыре метода кластеризации: DBSCAN, Affinity Propagation, k-means и иерархическую кластеризацию. По результатам сравнения рассмотренных методов автор делает вывод о том, что из рассмотренных методов наиболее подходящим для решения задачи является метод Affinity Propagation. При его использовании коэффициент силуэта получился наибольшим, что свидетельствует о большей согласованности данных кластеров. Этот метод также не требует больших временных затрат на поиски наиболее релевантных кластеров, в отличие от методов k-means и иерархической кластеризации.

В результате выполнения данной работы был создан программный продукт для сбора сообщений из сети «Твиттер», организована система хранения полученных сообщений, а также организована работа программного продукта в течении длительного времени. Были опробованы два вида векторного представления полученных текстовых сообщений и проведен сравнительный анализ известных методов кластеризации для выделения кластеров сообщений, объединенных одной темой.

В целом выпускная квалификационная работа Логачева М.М. 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. Структура работы соответствует логике исследования.

Таким образом, выпускная квалификационная работа Логачева Михаила Максимовича “Детектирование общественно значимых новостей в потоке сообщений” заслуживает оценки **отлично**.

Кандидат технических наук



Анфиногенов Сергей Олегович

31 мая 2018