

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.738.5

Жуковский Дмитрий Игоревич

ОТСЛЕЖИВАНИЕ ПЕРЕМЕЩЕНИЙ ЛЮДЕЙ
НА ОСНОВЕ ИХ АКТИВНОСТИ В СОЦИАЛЬНЫХ СЕТЯХ

АВТОРЕФЕРАТ

на соискание академической степени
магистра технических наук

По специальности 1-40 80 04 – Математическое моделирование, численные
методы и комплексы программ

Научный руководитель
кандидат технических наук, доцент
Стержанов М.В.

Минск 2015

ВВЕДЕНИЕ

Современные социальные сети предоставляют практически неисчерпаемый источник информации. Эта информация может носить разный характер, однако, важным отличительной особенностью социальных сетей является тесная связь пользователя и контента (записи, комментарии, отметки «мне нравится»).

В данный момент большая часть фотографий, публичных заметок и прочих социальных сигналов, публикуемых пользователями социальных сетей, содержит информацию о местоположении автора в момент публикации. Следовательно, анализируя данные социальных сетей можно узнать, какие бренды или продукты пользуются спросом, а какие теряют позиции.

С помощью анализа полученной и обработанной информации можно выяснить, какие общественные места посещаются целевыми группами пользователей в разное время суток, для определения целевой группы может использоваться фильтрация по половой принадлежности, возрастной группе, интересам и прочее.

Современные достижения в лексическом анализе текста позволяют выделить сообщения, направленные на поиск пользователем определенных товаров или услуг. Поэтому, проводя анализ тональности сообщений можно выяснить, какие события или факторы наибольшим образом влияют на настроение целевой группы пользователей.

Таким образом, актуальной и важной является разработка приложения для интеллектуального анализа данных (DataMining), находящейся в открытом доступе в социальных сетях.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Целью исследования является построение системы, способной собирать данные из социальных сетей в режиме реального времени, обрабатывать их и использовать результат работы для интеллектуального анализа данных.

Для достижения поставленной цели были выделены следующие задачи:

- обзор существующих социальных сетей, основных типов их контента;
- разработка системы сохранения данных из трех наиболее популярных на территории Республики Беларусь социальных сетей;
- построение системы интеллектуального анализа полученных данных.

Магистерская диссертация относится к активно развивающемуся перспективному направлению интеллектуального анализа данных. Исследование затрагивает технологические и алгоритмические основы.

Были разработаны приложения для сохранения данных социальных сетей в режиме реального времени с целью последующего анализа.

В ходе исследований был разработан ряд приложений для визуализации полученных данных.

Результаты исследования прошли апробацию на XII Белорусско–Российской научно–технической конференции «Техническиесредствазащитыинформации», которая состоялась в Минске в 2014 году.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Диссертация состоит из четырех глав, а также введения, заключения и приложения.

Во введении отражена постановка проблемы, а также возможное решение данной задачи путем построения системы интеллектуального анализа данных.

В первой главе производится обзор наиболее популярных социальных сетей Республики Беларусь. Приводятся основные типы публикуемого в них контента. Также показывается, каким образом публикуемый контент связывается с данными о местоположении пользователей в момент публикации.

Во второй главе производится анализ инструментальных средств, которые использовались при разработке приложения. Производится обзор решений для серверной и клиентской части приложения.

В третьей главе описывается реализация части приложения для сохранения данных об активности пользователей в социальных сетях. Приводится обзор API социальных систем и приемов для работы с ними. Также в третьей главе приводится описание слоя доступа к данным, структуры базы данных, описание основных сущностей и связей между ними.

В четвертой главе производится обзор основных компонентов для обработки полученных данных об активности пользователей социальных сетей. Приводится пример построения тепловой карты активности пользователей,

анализа тональности текстов и применения лексического анализа для поиска потенциальных клиентов для различных товаров и услуг.

По результатам проделанных исследований делается общий вывод.

Также к работе прилагается библиографический список.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Проведен тщательный обзор трех наиболее популярных на территории Беларуси социальных сетей с точки зрения анализа данных, содержащих информацию о локации пользователей. Были детально рассмотрены основные форматы контента и обосновано его использование для интеллектуального анализа данных (datamining) при решении практических задач.

2. Сформулированы и перечислены основные способы получения данных об активности пользователей социальных сетей. Разработан модуль анализа активности пользователей. Используя описанные в работе приемы можно отследить динамику взаимодействия пользователей социальной сети в режиме реального времени, что может найти непосредственное применение в статистических исследованиях, поиске потенциальных клиентов, мониторинге брендов, маркетинговых исследованиях и так далее.

3. На основе описанных в работе алгоритмов разработано веб-приложение, позволяющее аккумулировать данные социальных сетей и на их основе проводить анализ активности пользователей, оценку тональности сообщений, лексический анализ текста сообщений с целью их классификации. Приложение прошло проверку на реальных данных, что подтвердило возможность его применения для решения реальных практических задач.

Большая часть работы посвящена архитектуре разработанной системы, взаимодействию ее модулей, ключевых положений и наработок. Приведено описание основных компонентов и подходов, использованных при его реализации. Проведен анализ разработанных алгоритмов и их использования.

Значительное внимание было уделено оценке эффективности работы системы и алгоритма классификации, внедрение данных оценок в окончательный программный продукт, что позволяет производить мониторинг и администрирование системы с целью дальнейших потенциальных улучшений, а также для анализа поведения пользователей.

Рекомендации по практическому использованию результатов

1. Полученные результаты формируют теоретическую и практическую базу для разработки алгоритмов машинного обучения в области исследования активности пользователей социальных сетей. Они могут быть использованы для модернизации и дальнейшего развития существующих систем.

2. Разработанные алгоритмы лексического анализа текста сообщений пользователей для выявления потенциальных клиентов могут быть использованы в различных тематических веб-приложениях.

3. Результаты работы могут использоваться для интеллектуального анализа данных из социальных сетей, социологических и статистических исследований на основе этих данных.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1-А. Жуковский Д.И., "Мониторинг местоположения пользователей на основе их активности в социальных сетях / Д.И. Жуковский // Технические средства защиты информации: материалы XII Белорусско–Российской научно–технической конференции. - Минск БГУИР 2014. – с. 25.