

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

УДК 004.8

Свядыш
Дмитрий Алексеевич

Извлечение информации о происшествиях из публикаций в социальных сетях

АВТОРЕФЕРАТ

на соискание степени магистра информатики и вычислительной техники
по специальности 1-40 81 03 Искусственный интеллект

Научный руководитель
Сердюков Роман Евгеньевич
кандидат технических наук

Минск 2019

Нормоконтроль

Краткое введение

Современное общество так же, как и прежде, не защищено от чрезвычайных ситуаций и каких-либо происшествий. Важным является оперативное получение актуальной информации о таковых. Например, различные СМИ, а также государственные структуры, занимающиеся решением чрезвычайных ситуаций, крайне заинтересованы в получении таких данных.

Возможным решением данной задачи является создание автоматизированной системы, которая будет способна предоставлять актуальную информацию о происшествиях из публикаций в социальных сетях.

Общая характеристика работы

Актуальность темы исследования

Как правило, основным источником информации для СМИ являются пресс-релизы от служб спасения, однако данная ситуация имеет ряд недостатков. Во-первых, государственные службы не торопятся делиться известной им информацией со СМИ в течение некоторого времени. Во-вторых, пресс-релизы от служб, как правило, имеют задачу успокоить жителей, донести до них, что все под контролем, тем самым искажая реальную ситуацию. Данный факт может негативно сказываться на распространении достоверной информации. Также, для успешного функционирования, СМИ необходимо быть независимыми, освещать разные позиции и точки зрения. Использование одних лишь пресс-релизов и сообщений от государственных служб спасения может привести к однобокому освещению происшествий.

Возможным решением упомянутых выше проблем является создание автоматизированной системы реального времени, предоставляющей актуальную информацию о происшествиях из публикаций социальной сети. Решение описанных проблем представляет собой актуальную научную проблему, которая обладает значительной новизной.

Цель и задачи исследования

Цель диссертации состоит в разработке системы извлечения информации о происшествиях из публикаций социальной сети для решения задач, с которыми сталкиваются СМИ.

В соответствии с целью в диссертации решаются следующие *задачи*:

- исследовать аналогичные системы, на основе результатов исследования выявить требования к разрабатываемой системе;
- спроектировать предложенную систему;
- разработать и протестировать спроектированную систему на реальных данных, провести анализ полученных результатов, определить перспективы разработки систем для извлечения информации о происшествиях.

Опубликование результатов диссертации

Некоторые результаты исследования были опубликованы на двух научных конференциях аспирантов, магистрантов и студентов БГУИР.

Структура и объем диссертации

Структура диссертационной работы обусловлена целью, задачами и логикой исследования. Работа состоит из введения, трех глав, заключения и библиографического списка. Общий объем диссертации – 45 страниц. Работа содержит 23 иллюстрации и 4 таблицы. Библиографический список включает 33 наименования.

Краткое содержание работы

Во **введении** рассмотрено современное состояние проблемы получения актуальной информации о происшествиях для СМИ, а также дается обоснование актуальности темы диссертационной работы.

В **первой главе** исследованы существующие решения классификации и извлечения информации из текстов, написанных на естественном языке. Были подробно изучены методы машинного обучения для обработки естественного языка, определены слабые и сильные стороны каждого из них. В результате анализа области исследования были сформулированы требования к разрабатываемой системе, которые призваны решить описанные проблемы в работе СМИ.

Во **второй главе** была спроектирована система обнаружения и извлечения информации о происшествиях из публикаций в социальных сетях. В соответствии с требованиями к разрабатываемой системе были выделены основные функции системы. Задачи и принципы работы основных компонентов системы были подробно описаны.

В **третьей главе** отображены подробности реализации разрабатываемой системы, ее подробное описание и сценарии использования. Также был проведен ряд экспериментов на реальных данных, по результатам которых была показана жизнеспособность предложенной системы, а также в ходе тестирования был улучшен компонент, ответственный за классификацию публикаций. Помимо этого, были выявлены недостатки разработанной системы и предложены возможные улучшения.

Заключение

В ходе данной работы были рассмотрены проблемы в работе СМИ, предложена система, призванная решить их. Для формулирования требований к разрабатываемой системе были изучены существующие аналоги, а также методы машинного обучения для обработки естественного языка.

Была спроектирована и разработана система обнаружения и извлечения информации о происшествиях из публикаций в социальных сетях. Также был проведен ряд экспериментов, по результатам которых была показана жизнеспособность предложенной системы. В дальнейших исследованиях планируется улучшать точность нахождения публикаций, связанных с происшествиями, а также расширить возможности системы для удобной работы с обнаруженными происшествиями.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Свядыш Д.А. Извлечение информации о происшествиях из публикаций в социальных сетях / 54-ая научная конференция аспирантов, магистрантов и студентов БГУИР. – 2018. – в печати
2. Свядыш Д.А. Анализ существующих средств извлечения структурированных данных из текста на естественном языке / 55-ая научная конференция аспирантов, магистрантов и студентов БГУИР. – 2019. – в печати