

АННОТАЦИЯ

к диссертационной работе докторанта PhD Болатбек Миланы Асланбекқызы по специальности «Системы информационной безопасности» 6D100200 на тему: «Разработка и исследование моделей семантического анализа в веб-ресурсах для определения экстремистской направленности в тексте»

Актуальность темы исследования. В настоящее время информационно-коммуникационная сеть Интернет стала неотъемлемой частью жизни человечества. Люди используют социальные сети «Твиттер», «ВКонтакте», «Facebook» и т. д. для глобального общения, обмена мнениями, получения знаний и т. д. Активное участие не только отдельных пользователей, но и информационных организаций во всем мировом пространстве определяет необходимость разработки мероприятий по обеспечению национальной безопасности, соответствующих современным тенденциям развития информационно-коммуникационных технологий, в частности организации мероприятий, касающихся противодействия усилению идей экстремизма и терроризма. Экстремистские организации используют новые информационные технологии с целью привлечения новых членов в группы, планирования и выполнения экстремистской деятельности, проведения обучающей работы, обмена конфиденциальной информацией в управлении и координации социально-опасной деятельности, поиска источников финансирования для осуществления экстремистской деятельности, создания закрытых сайтов с целью целенаправленной идеологической пропаганды пользователей, в том числе молодежи и т.п. в рациональном выполнении действий. Проблема носит всемирный характер и является очень актуальной и для Республики Казахстан как одного из ключевых участников глобального политического процесса.

Информационное пространство Интернет состоит из различных ресурсов. При этом подавляющее большинство из них не являются средствами массовой информации, в результате чего применение норм законодательства о средствах массовой информации становится невозможным. Для противодействия распространению идей экстремизма в настоящее время правоохранительными органами используются нормы уголовного законодательства от 18 февраля 2005 года № 31 «О противодействии экстремизму», «Об утверждении государственной программы по противодействию религиозному экстремизму и терроризму в Республике Казахстан на 2018-2022 годы», Постановление Правительства Республики Казахстан № 124 от 15 марта 2018 года.

В настоящее время в национальный список запрещенных террористических формирований на территории Казахстана включены 22 организации.

В ноябре 2020 года Президент РК Токаев К. К. призвал государства-члены ШОС создать Центр информационной безопасности, направленный на борьбу с сепаратизмом, терроризмом и экстремизмом.

К сожалению, в последние годы казахстанцы также входят в число экстремистских организаций. По данным Комитета национальной безопасности (КНБ), с начала войны в Сирии и Ираке на Ближний Восток выехали около 800 граждан Казахстана. Многие из них-дети. По состоянию на июль 2018 года в международных террористических организациях в Сирии и Ираке находятся 120 казахстанских мужчин, более 250 женщин и 500 несовершеннолетних детей.

По поручению Елбасы Нурсултана Назарбаева с января 2019 года в рамках операции "Жусан" начали возвращать граждан Сирии и Ирака. В ответе комитета от ноября 2019 года говорится, что из Сирии и Ирака в Казахстан возвращено 277 взрослых (57 мужчин, 220 женщин) и 547 несовершеннолетних детей, в Сирии находятся еще более 90 граждан.

Данные вопросы можно рассматривать как угрозу национальной безопасности Республики Казахстан. Ситуация в сфере противодействия экстремистской деятельности в интернете остается сложной, что требует проведения научных исследований и осуществления комплекса эффективных и своевременных мер, направленных на выявление, предупреждение и пресечение любых проявлений экстремизма.

Гиганты Google, Facebook и Twitter пообещали использовать технологию искусственного интеллекта (AI) для быстрого обнаружения и уничтожения террористического контента в интернете.

У IBM есть разработка Watson, которая может анализировать все данные из вышеупомянутых социальных сетей. В России IT-автор Платона создает систему мониторинга социальных сетей и прогнозирования рисков. Правительство Германии объявило о создании нового подразделения кибербезопасности под названием ZITiS для борьбы с террористами в Интернете после террористических актов. Таких систем в Казахстане пока нет. По этой причине проведение анализа веб-ресурсов в Интернете, автоматизация своевременного выявления текстов экстремистского содержания являются наиболее актуальными для организаций по борьбе с экстремизмом.

Развитие социальных сетей способствует быстрому распространению групп сторонников насилия, пропагандистов экстремизма и радикализма. Сайты микроблогов, работы по выявлению экстремистского контента в группах социальных сетей являются сложной и развивающейся областью исследований. Проблема экстремизма исследуется в отечественной и зарубежной научной литературе с XX века по сегодняшний день. На сегодняшний день экстремизм является объектом исследования различных научных работ.

В настоящее время подавляющее большинство работ, связанных с выявлением экстремистских текстов на веб-ресурсах, написаны для английского языка. Ученые эффективно используют методы машинного и глубокого обучения для выявления экстремистских направлений в веб-ресурсах. В последнее время на веб-ресурсах создаются различные системы для выявления текстов, содержащих экстремистский контент, написанный на немецком, русском, арабском языках.

Проведен ряд научных исследований, посвященных выявлению и анализу текстов криминального характера для казахского языка. В частности, можно привести труды ученых научно-исследовательского института «Искусственный интеллект» при Евразийском национальном университете имени Л. Н. Гумилева под руководством Шарипбая А. А., так же в РГП на ПХВ «Институт информационных и вычислительных технологий» Мамырбаева О. Ж. В указанных работах описывается метод, основанный на правилах, с использованием словаря для анализа настроений текстов на казахском языке, связанных с террористическими угрозами. В нем представлен обзор методов полярного анализа, парсер, анализирующий страницы по содержанию ключевых слов в базе данных, морфологический, синтаксический и сентиментальный анализ текстов на казахском языке. Создана база онтологических знаний, используемых для разработки программного комплекса для идентификации пользователя социальных сетей при обеспечении информационной безопасности, мониторинге и предупреждении угроз.

С учетом вышеизложенных выводов можно сделать вывод, что **наиболее актуальной** является задача построения моделей для выявления текстов экстремистской направленности на казахском языке на веб-ресурсах.

Цель диссертационной работы. Комплексное исследование и построение модели и метода семантического анализа для выявления экстремистских текстов на казахском языке на веб-ресурсах. Исследование точности определения экстремистских текстов с помощью построенной модели и метода.

Задачи исследования. Для достижения поставленной цели предусматривается выполнение следующих задач:

- 1) создание корпуса экстремистских текстов на казахском языке для обучения и тестирования методов машинного и глубокого обучения для выявления экстремистских текстов на казахском языке на веб-ресурсах;
- 2) Создание модели семантического анализа для выявления экстремистских текстов на казахском языке на веб-ресурсах;
- 3) Создание и исследование методов для выявления экстремистских текстов на казахском языке на веб-ресурсах;
- 4) составление списка экстремистских ключевых слов на казахском языке;
- 5) создание и тестирование программного обеспечения для выявления экстремистских текстов на казахском языке на веб-ресурсах на основе разработанных моделей и методов.

Объект исследования. Выявление текстов экстремистской направленности на Веб-ресурсах.

Предмет исследования. Методы машинного и глубокого обучения для выявления текстов экстремистской направленности на веб-ресурсах.

Метод исследования. В качестве метода исследования были использованы методы машинного и глубокого обучения, методы классификации текстов, методы обработки естественного языка, нейронные сети, методы анализа социальных сетей, методы статистической обработки, методы системного анализа.

Научная новизна работы:

1) впервые создан корпус экстремистских текстов на казахском языке для обучения и тестирования методов машинного обучения для выявления экстремистских текстов на казахском языке;

2) впервые с учетом особенностей казахского языка построена модель семантического анализа, отличающаяся применением метода TF-IDF к биграммам, предварительно примененным алгоритмом стемминга в слой внедрения слов сети LSTM и повышающая точность определения экстремистских текстов;

3) разработан метод формирования набора признаков, который основывается на комбинациях N-грамм и методов ввода слова и повышает качество классификации экстремистских текстов;

4) впервые создан список экстремистских ключевых слов на казахском языке.

Теоретическая значимость работы. Теоретическая значимость диссертационной работы основана на совокупности знаний в области методов и алгоритмов определения экстремистских действий и организаций. Полученные фундаментальные результаты могут быть использованы мировым научным сообществом.

Практическая значимость работы. Прикладные результаты в виде метода, авторских свидетельств могут быть использованы уполномоченными органами по обеспечению информационной безопасности, критической инфраструктуры, борьбе с интернет-экстремизмом.

Основное положение выносимое на защиту. Создан корпус экстремистских текстов на казахском языке для обучения и тестирования методов машинного обучения. Разработана модель семантического анализа для выявления текстов экстремистской направленности на казахском языке на веб-ресурсах. Установлено, что в модели LSTM можно повысить точность задания определения экстремистских текстов на казахском языке на веб-ресурсах с помощью применения метода ввода слов TF-IDF на биграммы, предварительно примененным алгоритмом стемминга.

Степень надежности и результаты апробации. Достоверность и обоснованность результатов исследования подкрепляются обоснованной ответственностью постановки задач, экспертизой критериев и состояния исследований в данной области, большим количеством проведенных экспериментов, а также их успешным внедрением в практику. Результаты диссертации докладывались и обсуждались на следующих научно-методических конференциях:

1) Mussiraliyeva Sh., Bolatbek M., Omarov B., Bagitova K. Detection Of Extremist Ideation On Social Media Using Machine Learning Techniques // 12th International Conference on Computational Collective Intelligence. – Vietnam, 2020. – P.743-752, https://doi.org/10.1007/978-3-030-63007-2_58

2) Mussiraliyeva Sh., Bolatbek M., Omarov B., Medetbek Zh., Baispay G., Ospanov R. On Detecting Online Radicalization and Extremism Using Natural Language Processing // 21st International Arab Conference on Information Technology (ACIT'2020). – Egypt, 2020, DOI: 10.1109/ACIT50332.2020.9300086

3) Mussiraliyeva Sh., Omarov B., Bolatbek M., Ospanov R., Baispay G., Medetbek Zh., Yeltay Zh. Applying Deep Learning for Extremism Detection // International Conference on Advanced Informatics for Computing Research. – Singapore, 2021. – P.597-605. https://doi.org/10.1007/978-981-16-3660-8_56

4) Mussiraliyeva Sh., Bolatbek M., Omarov B., Bagitova K., Alimzhanova Zh. Bigram based Deep Neural Network for Extremism Detection in Online User Generated Contents in the Kazakh Language // International Conference on Computational Collective Intelligence. – Greece, 2021. – P.559-570. https://doi.org/10.1007/978-3-030-88113-9_45

5) Болатбек М.А., Создание словаря экстремистских слов для казахского языка, Международная научная конференция студентов и молодых ученых «ФАРАБИ ЭЛЕМИ», Казахстан, Алматы, 2018

6) Мусиралиева Ш.Ж., Болатбек М.А., Әлеуметтік желідегі экстремистік мәтіндерді жіктеу дәлдігін грамматикалық қателерді анықтау және түзету арқылы арттыру, Международная научно-практическая конференция "Актуальные проблемы информационной безопасности в Казахстане, 2020

7) Болатбек М.А., Экстремизм түсінігі. Экстремистік мәтіндерді анықтауға арналған белгілер жинағына шолу, Международная научная конференция студентов и молодых ученых «ФАРАБИ ӘЛЕМІ», Казахстан, Алматы, 2020

8) Болатбек М.А., Экстремистік мәтіндерді сентимент талдау арқылы анықтау, Международная научная конференция студентов и молодых ученых «ФАРАБИ ӘЛЕМІ», Казахстан, Алматы, 2020

9) Байдулла А.М., Мусиралиева Ш.Ж., Болатбек М.А. Экстремистік топтарды анықтау және талдау // Матер. Междунар. научн. конф. студентов и молодых ученых «Фараби әлемі». – Алматы: Қазақ университеті, 2021. – С. 74.

10) Мусиралиева Ш.Ж., Болатбек М.А. Веб-ресурстардағы экстремистік мәтіндерді анықтаудың семантикалық үлгілерін құру және зерттеу // Матер. Междунар. научн. конф. студентов и молодых ученых «Фараби әлемі». – Алматы: Қазақ университеті, 2021. – С. 77.

11) Маден М.Т., Мусиралиева Ш.Ж., Болатбек М.А. Онлайн ортада экстремизмнің лингвистикалық маркерлерін анықтау // Матер. Междунар. научн. конф. студентов и молодых ученых «Фараби әлемі». – Алматы: Қазақ университеті, 2021. – С. 101.

12) Шәріпбекова С.Е., Мусиралиева Ш.Ж., Болатбек М.А. Қазақ тіліндегі оң қанатты экстремизмді анықтау үшін веб-контентті жинауға арналған бағдарламалық модуль әзірлеу // Матер. Междунар. научн. конф. студентов и молодых ученых «Фараби әлемі». – Алматы: Қазақ университеті, 2021. – С. 118.

13) Ынтықбай Б.Н., Мусиралиева Ш.Ж., Болатбек М.А. Әлеуметтік желілердегі қауіпсіздік пен конфиденциалдықты машиналық оқыту тәсілдерін қолдану арқылы талдау // Матер. Междунар. научн. конф. студентов и молодых ученых «Фараби әлемі». – Алматы: Қазақ университеті, 2021. – С. 119.

14) Мусиралиева Ш.Ж., Омаров Б.С., Болатбек М.А., Жастай Е. Веб-ресурстардағы қазақ тіліндегі экстремистік сипаттағы мәтіндерді анықтау // Матер. Междунар. научн. конф. в области информационных технологий, посвященной 75-летию профессора У.А.Тукеева. – Алматы, 2021. – С. 98-104.

Личный вклад соискателя. Соискатель решила поставленные задачи диссертационной работы. Разработана семантическая модель и метод выявления текстов экстремистской направленности на казахском языке на веб-ресурсах. Создан корпус экстремистских текстов на казахском языке для обучения и тестирования алгоритмов машинного обучения. Проведены эксперименты с целью определения точности разработанной модели и метода. Создан список экстремистских ключевых слов на казахском языке.

Связь темы диссертации с планами научно-исследовательских работ. Данная работа выполнена по заказу Министерства цифрового развития, инноваций и аэрокосмической промышленности Республики Казахстан в рамках научно-исследовательских работ проекта «Разработка моделей, алгоритмов семантического анализа веб-контента для определения экстремистской направленности и создание инструментария кибер-криминалистики», ИРН АР06851248. Приоритетное направление: национальная безопасность и оборона. Специализированное направление: обеспечение информационной безопасности. Исследования, проведенные по диссертационной работе, включены в отчет данного проекта за 2020-2021 годы.

Структура и объем работы. Диссертационная работа состоит из введения, 4 глав, заключения, списка литературы и приложений. Полный объем диссертации: 112 страниц машинописного текста, включающего 71 рисунок, 14 таблиц, 137 списков использованной литературы и 4 приложения.

Публикации результатов. В ходе выполнения научно-исследовательских работ было написано 19 научных работ. Из них 1 статья опубликована в журнале «Computers, Materials & Continua», индексируемом на базе Scopus, 4 статьи в изданиях, рекомендованных комитетом по

контролю в сфере образования и науки Министерства образования и науки Республики Казахстан, 14 статей в сборниках международных научно-практических конференций.

Во введении дается обоснование актуальности темы диссертационной работы. Сформулированы цель, объект и предмет научно-исследовательской работы. Вместе с тем, показана научная новизна и практическая значимость. Приведены сведения об апробации и публикации результатов исследовательских работ.

В первом разделе приводится понятие экстремизма, классификация экстремизма. Приводится обзор глобальных экстремистских ресурсов. Дается определение противодействию экстремизму, рассматриваются международные проекты противодействия экстремистской деятельности.

Во втором разделе рассматривается задача создания корпуса, необходимого для построения семантической модели обнаружения экстремистских данных на веб-ресурсах. Дается характеристика работы программного обеспечения для сбора и анализа веб-контента для определения экстремистской направленности. Приводится статистика по составленному корпусу и приводятся особенности текстов в категориях корпусов. Кроме того, для данных корпуса предусмотрено выполнение задания морфологического анализа и сформирован список ключевых слов. В результате данного раздела впервые создан корпус экстремистских текстов на казахском языке для обучения и тестирования методов машинного обучения для выявления экстремистских текстов на казахском языке.

В третьем разделе рассматривается задача построения семантической модели для выявления экстремистских текстов на веб-ресурсах и анализируется архитектура построенной модели и дается характеристика ее модулей. С целью определения высокоточного метода выявления экстремистских текстов на казахском языке на веб-ресурсах проводятся эксперименты с использованием различных методов ввода слов и n-грамм. В результате проведенных исследований был разработан метод формирования набора признаков, основанный на комбинациях N-грамм и методов ввода слова, который повышает качество классификации экстремистских текстов и впервые с учетом особенностей казахского языка разработана модель семантического анализа, отличающаяся применением метода TF-IDF к биграммам, предварительно примененным алгоритмом стемминга в слой ввода слов сети LSTM и повышающая точность определения экстремистских текстов.

В четвертом разделе проводится сравнительный анализ различных методов машинного обучения и предлагаемого метода для выявления экстремистских текстов на казахском языке на веб-ресурсах.

В заключении формулируются основные выводы и результаты работы.