

УДК 004.912

**ОПРЕДЕЛЕНИЕ ИНТЕРНЕТ РЕСУРСОВ СОДЕРЖАЩИХ ЗАПРЕЩЕННЫЙ
КОНТЕНТ**

Аубакирова Салтанат Маратовна

aubakirovasaltanat@yandex.kz

Магистрант 2-го курса специальности «Системы информационной безопасности»

ЕНУ им. Л.Н.Гумилева, Нур-Султан, Казахстан

Научный руководитель - Г.Т. Бекманова

Введение

В соответствии с «концепцией информационной безопасности» Республики Казахстан, понятие «информационная безопасность» состоит из двух взаимосвязанных технического и социально-политического аспектов. Технический аспект предполагает обеспечение защиты национальных информационных ресурсов, информационных систем, информационно-телекоммуникационной инфраструктуры, обеспечения целостности, конфиденциальности и доступности информации, защиты от несанкционированного доступа, использования, раскрытия, нарушения, обучения, проверки, записи или уничтожения информации. Социально-политический аспект представляет собой защиту национального информационного пространства и систем распространения массовой информации от негативного информационного и организованного воздействия, которое может нанести ущерб национальным интересам Республики Казахстан [1]. В этой связи социально-политический аспект подразумевает защиту от негативной информации, которая может нанести ущерб национальным интересам Республики Казахстан.

Террористические организации умело используют интернет в своих собственных целях, оказывая психологическое воздействие на граждан посредством агитации и дальнейшего распространения собственных идеологий, тем самым привлекая больше новобранцев в свои ряды.

Согласно данным методических рекомендаций НАТ РФ (Национальный антитеррористический комитет) в 1998 году террористические организации вели активную деятельность используя лишь 12 сайтов, к 2005 году их количество увеличилось до 4800, а на 2013 год их насчитывалось более 10000 [2]. На сегодняшний день нет точной количественной информации по этому вопросу, так как ежедневно создается огромное количество подобных сайтов.

Интернет - самый масштабный ресурс, обеспечивающий широкий спектр информации. При упоминании понятия «информации», отмечается её тема и целевая аудитория, в этой связи информацию в интернете можно разделить на две основные категории: разрешенный и запрещенный контент. Последний, в свою очередь, является предметом исследования.

В целом проблематика определения запрещенного контента в сети сводится к сентимент-анализу (анализ тональности). Анализ тональности – это ряд методов, приемов и инструментов, описывающих выявление и извлечение субъективной информации, такой как мнение и отношение из текста [3]. Сентимент-анализ подразумевает оценку полярности мнений, то есть, имеет ли высказывание положительную, нейтральную или отрицательную эмоциональную окраску.

Первые работы по определению запрещенного контента посредством сентимент-анализа приходятся на начало 2000-х годов. Наиболее значимые из них являются следующие: (Tom De Smedt, Guy De Pauw and Pieter Van Ostaeyen, 2018) «Автоматическое онлайн обнаружение джихадистских выражений», в данной работе описывается работа системы, которая обнаруживает запрещенные высказывания с точностью более 80%, используя приемы обработки естественного языка и машинное обучение. Следующая статья рассматривает использование методов интеллектуального анализа данных для обнаружения запрещенной террористической деятельности в сети (Mahesh S., Mahesh T.R., M.Vinayababu, 2010).

В статьях по исследованию основных подходов по сентимент-анализу представлены сравнительные анализы существующих методов, с описанием преимуществ и недостатков каждого из них (Devika M.D., Sunitha C., Amal Ganesha, 2016; Salma Faiz Alasmari, Mohammed Dahab, 2017). Несмотря на то, что на данный момент ни один из методов не предоставляет 100% точный результат обработки текстовых данных, сентимент-анализ активно используется для обнаружения определенной информации.

Предметом исследования является запрещенный контент в Интернете. Контент – вид информации, являющейся содержанием информационного продукта (текст, аудио, видео,

фото), то есть, контент – это внутренняя составляющая информационного продукта, которая определяет его качество и репутацию. По способам передачи информации различают четыре различных вида контента:

- Текстовый контент;
- Аудиоконтент;
- Видеоконтент;
- Изображение.

Запрещенный контент подразумевает любой вид информации, несущий в себе юридически запрещенный на государственном уровне смысл. В большинстве случаев он представляет собой тексты, аудио и видеоматериалы, пропагандирующие индивидов на совершение действий, являющимися незаконными и противоречащими интересам государства [4]. Запрещенный контент включает в себя, но не ограничивается:

- Информация, пропагандирующая и/или оправдывающая насилие в отношении личности или группы, и представляющая опасность для этих индивидов;
- Информация порнографического характера;
- Информация, связанная с продажей, пропагандой употребления наркотических веществ;
- Информация, связанная с продажей, пропагандой чрезмерного употребления алкогольных продуктов;
- Информация, связанная с продажей, пропагандой чрезмерного употребления табачных изделий;
- Информация, связанная с нарушением программ и компьютерных систем;
- Информация с элементами насилия;
- Информация, связанная с продажей оружия;
- Информация, связанная с пропагандой, агитацией терроризма и экстремизма и распространение данной идеологии, публичные призывы к выполнению террористического акта [5].

Для данного исследования был выбран метод сентимент-анализа на основе базы знаний, состоящей из словаря (база данных). Объект исследования – запрещенный террористический, экстремистский текстовый контент. Основная идея данной статьи – построение уникальной трехязычной базы знаний, адаптированной и идентичной для всех трех языков (русский, казахский, английский), служащей для определения интернет ресурсов, содержащих запрещенный террористический текстовый контент. В связи с отсутствием определенной утвержденной базы знаний для подобных целей, был использован метод «ручного» сбора информации на примере реальных, существующих интернет ресурсов содержащих и распространяющих запрещенный контент. В качестве примера использовались персональные профили пользователей социальной сети «ВКонтакте» и «Facebook», зарубежные сайты.

Несмотря на отсутствие готовых баз знаний под определенные тематические области, существует ряд баз данных, специально размеченных с учётом эмоциональной составляющей. Такие как SentiWordNet, WordNet-Affect и SenticNet. Данные базы знаний представлены на русском и английском языках. Вышеперечисленные базы знаний использовались в качестве образца при разработке собственной базы данных, с дополнением в виде казахского языка.

Как было указано ранее, для исследования были выбраны социальные сети «ВКонтакте» и «Facebook» и также отдельные зарубежные интернет ресурсы распространяющие террористический контент. Для начала были напрямую рассмотрены отдельные профили пользователей социальных сетей, упоминающие запрещенные слова из списка, предоставленным Американским департаментом по национальной безопасности в 2011 году, использующегося для анализа сетей на запрещенный контент [6].

В ходе исследования подобных страниц была собрана информация о основных путях распространения материала теороричтического характера на социальных площадках. В

основном ими являются текстовые материалы, так называемые «записи на стене; пост» [7], которые пользователи выкладывают на своих страницах. Было рассмотрено вручную более 50 подобных страниц для непосредственного составления базы данных на данной основе. Основные языки на которых публиковался материал – английский, русский, казахский (Рис. 1) (Рис. 2)

Таблица пользователей "Вконтакте"



Рисунок 1 – Информация по пользователям «Вконтакте»

Таблица пользователей "Facebook"



Рисунок 2 – Информация по пользователям «Facebook»

На зарубежных ресурсах, в большинстве, информация была представлена в виде видео материалов агитационного и насильного характера, имеющих краткое описание. (Рис. 3) (Рис. 4)

ISIS Execution By Beheading, Gunfire & Crushing Head With A Rock - Video



Рисунок 3 – Пример видеоматериала террористического характера



Рисунок 4 – Пример видеоматериала террористического характера

На основе проведенного анализа была создана тестовая база данных, содержащая наиболее часто употребляемые в запрещенном контексте термины, слова, словосочетания (разбитые по ключевым словам), которая позже была преобразована в базу знаний.

Согласно стандарту, ISO/IEC/IEEE 24765-2010 [8] понятие базы знаний представляет следующее описание, это база данных, содержащая правила вывода и информацию о человеческом опыте и знаниях в некоторой предметной области. Подобные базы чаще используются при поиске и экстракте информации. Они включают в себя основные элементы (термины), отношения (частное – целое; объект – свойства; причина – следствие; функциональная схожесть и т.д.) (Рис. 5) (Рис. 6).

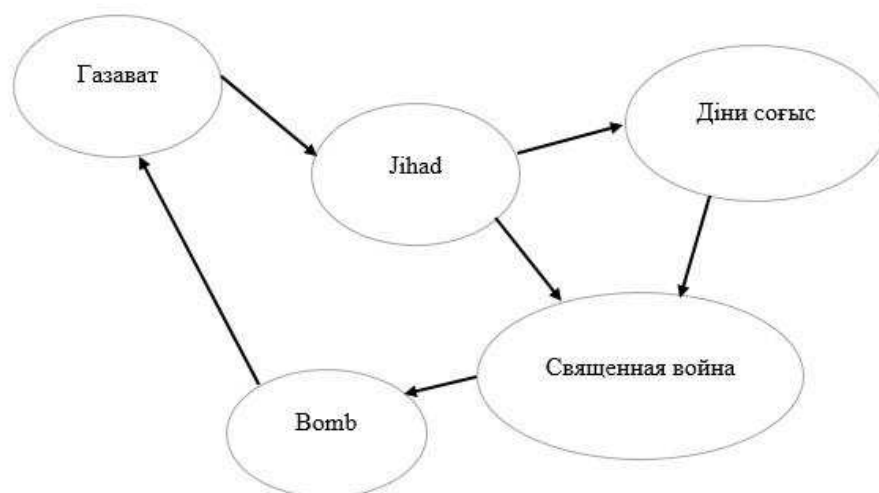


Рисунок 5 – Связи базы данных



Рисунок 6 – Связи базы данных

База знаний построена при учете следующих отношений: гиперонимов, меронимов, антонимов, синонимов, холонимов и омонимов.

Как представлено в рис.5, уникальность данной базы знаний в том, что она абсолютно идентична для трех языков, соответственно информация находится при любом сочетании слов. В свою очередь, это очень актуально для пользователей Казахстана, ввиду постоянного смешанного использования языков в речи.

Результативность созданной базы данных была проверена на основе анализа интернет-ресурсов на признаки содержания запрещенного контента посредством парсинга.

Парсинг это – синтаксический анализ текста. (Li Dong, Furu Wei, Shujie Liu, Ming Zhou, Ke Xu, 2015). В данном случае, программа парсер проверяет текст размещенный на интернет ресурсах на совпадения с базой знаний.

Принцип работы парсера следующий: программа переходит по всем внутренним ссылкам на заранее указанном ресурсе для проверки, исходный результат записывается в файл формата .txt, с указанными ссылками ключевыми словами и количеством совпадений по ним.

Выводы

Данная версия базы данных удобна в использовании для учебных, экспериментальных действий на начальном уровне. Собранные данные идентичны на всех трех языках (казахский, русский и английский), интерфейс программы также представлен в двух языках (русский, казахский). Однако, для более глубокого анализа планируется расширить базу знаний для глубокого семантического анализа текстов и, соответственно, точных результатов.

Список использованных источников

1. Концепция информационной безопасности Республики Казахстан до 2016 года [Электронный ресурс]. URL: <https://zonakz.net/2006/10/20/концепция-информационной-безопаснос/>
2. Методические рекомендации по совершенствованию пропагандистской работы в сфере противодействия распространению идеологии терроризма в субъектах российской федерации [Электронный ресурс], URL: <http://rk.gov.ru/file/metod2013.pdf>
3. B. Liu, “Handbook Chapter: Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing,” Handbook of Natural Language Processing. Marcel Dekker, Inc. New York, NY, USA, 2009
4. Prohibited Internet Content (white paper) by eSecureData Inc.
5. Определение запрещенного контента [Электронный ресурс], URL: <http://www.securitylab.ru/news/tags/%E7%E0%EF%F0%E5%F9%E5%ED%ED%FB%E9+%EA%EE%ED%F2%E5%ED%F2/>
6. Analyst’s Desktop Binder, U.S. Department of Homeland Security [Электронный ресурс], URL: <https://epic.org/foia/epic-v-dhs-media-monitoring/Analyst-Desktop-Binder-REDACTED.pdf>
7. Глоссарий интернет маркетинга [Электронный ресурс], URL: <https://glossary-internet.ru/terms/%CF/3556/>
8. Стандарт ISO/IEC/IEEE 24765-2010 [Электронный ресурс], URL: <https://www.smaele.nl/documents/iso/ISO-24765-2010.pdf>