

**М. И. Ананьева¹, М. В. Кобозева¹, Ф. Н. Соловьев²
И. В. Поляков³, А. М. Чеповский^{2,3}**

¹ ИСА ФИЦ «Информатика и управление» РАН
пр. 60-летия Октября, 9, Москва, 117312, Россия

² Московский политехнический университет
ул. Б. Семеновская, 38, Москва, 107023, Россия

³ Национальный исследовательский университет Высшая школа экономики
ул. Мясницкая, 20, Москва, 101000, Россия

achepovskiy@hse.ru

О ПРОБЛЕМЕ ВЫЯВЛЕНИЯ ЭКСТРЕМИСТСКОЙ НАПРАВЛЕННОСТИ В ТЕКСТАХ *

Рассматриваются методы, используемые для обнаружения экстремистских текстов из Интернета. Дальнейшее исследование необходимо, чтобы найти новые методы классификации текстов и дифференцирующие признаки для классификации текстов экстремистской тематики.

Ключевые слова: классификация текстов, тексты экстремистской направленности.

Задача экспертизы текстов

Противодействие распространению агрессивной информации в глобальной сети является актуальной проблемой общества и государственных органов, которая решается в частности фильтрацией нежелательных ресурсов Интернета [11]. Необходимым условием такой фильтрации является классификация содержания сайтов и документов информационного потока [11, 28]. Поэтому актуальной проблемой информационных технологий является классификация и рубрикация по тематической, психолингвистической направленности текстов на естественных языках [7, 15, 16, 28]. В частности, в [15, 16] обсуждаются актуальные проблемы судебной лингвистики, связанные с изучением «дискурса вражды», как проявления экстремистской направленности текстов и речевой агрессии; дается характеристика различных видов речевой враждебности.

В [2] проводится анализ текстов с экстремистским содержанием, на основе которого выводятся психологические критерии, по которым эксперт должен оценивать текст. В [27] обсуждается экспертиза, включающая несколько видов анализа. Предметно-содержательный анализ текста с целью экспертной оценки предполагает введение изучаемого преступного действия в понятийное поле психолога-эксперта. Социально-психологический подход является значимым при оценке материалов экстремистской направленности и отражает основные

* Работа выполнена при поддержке РФФИ, грант № 16-29-09546.

положения теории социально-психологического воздействия на различные группы. К сожалению, работы типа [2, 7, 15, 16, 27] не формулируют критерии выявления экстремистской направленности для автоматического анализа текстов.

Популярные в академической среде методы нейронных сетей и опорных векторов требуют большие обучающие выборки, что, во-первых, ограничивает работу информационных систем в режиме реального времени при решении задач тематической направленности текстов [3, 12, 13, 28], а во-вторых, для русского языка трудно собирать коллекции материалов в условиях отсутствия таких текстов в открытом доступе по закону РФ.

В любом случае создание соответствующих коллекций является сложнейшей актуальной проблемой выявления противоправных текстов и словарей соответствующей лексики [1, 14].

В [30] представлены примеры сбора, анализа и визуализации террористических материалов, находящихся в открытом доступе, с помощью так называемого “теневого Интернета” – сегмента сети, куда можно попасть лишь с помощью специального ПО, сохраняя там полную анонимность. Для исследования авторы взяли списки террористических группировок (664 организации) и их сайтов из правительственных источников США, и скачали их содержимое (3,6 миллионов веб-страниц) на английском, арабском и испанском языках. Готовый корпус экстремистских текстов на английском языке описывается в [31]. Все тексты были написаны на арабском языке и позже переведены на английский. Корпус имеет разноплановую разметку (синтаксическую, семантическую, анафорическую разметку, также размечены временные маркеры и события), которая проводилась автоматически, а затем проверялась вручную.

В [1] заявлено о создании корпуса текстов экстремистской направленности. Собранные тексты с противоправным содержанием разделены на семь категорий, включающие такие категории, как «терроризм» (объединяет тексты, пропагандирующие запрещенные в РФ организации), «религиозная ненависть» (содержит тексты, призывающие к активным действиям против людей других религий), «национализм» (ненависть к представителям других национальностей), «агрессия и призывы к беспорядкам» (призывают к участию в несанкционированных митингах и беспорядках, насильственному свержению власти). В [6] описывается создание корпуса сообщений пермского сегмента социальной сети “ВКонтакте”, содержащих признаки этнической агрессии.

Оценки эмоционального воздействия

В статье [6] анализируется коллекция сообщений пользователей социальной сети, указавших город Пермь местом жительства. По результатам исследований был составлен ранжированный словарь терминов, употребление которых в тексте может свидетельствовать о наличии агрессии. Составление словаря-гlossария терминов этнической агрессии рассматривается как важное методологическое действие, которое позволяет оптимизировать представления о данной предметной области. Предлагается строить иерархические структуры понятий для описания этнической агрессии. Но, по сути, получен словарь лексических терминов, который предлагается использовать для поиска ресурсов, содержащих феномены этнической агрессии в информационных ресурсах различной природы.

В [24] представлен анализ особенностей экстремистских текстов, размещаемых в Интернете на русском языке, для целей фильтрации нежелательного контента. В качестве языковых признаков экстремизма рассматриваются базы данных наборов лемм слов, которые обрабатываются при помощи грамматического словаря.

В [26] по результатам статистического анализа корпуса прецедентных текстов на исламскую тематику извлекается множество ключевых понятий. Практически составляются словари сигнальных слов, наличие которых в тексте позволяет экспертам сделать выводы о включении текста определенную категорию. Предлагается определять принадлежность текста к интересующей исламской тематике по вхождению в эти тексты сигнальных слов.

Экстремистская направленность текста на естественном языке определяется во многом эмоциональной компонентой текста. Поэтому важны методы, связанные с определением так называемой «тональности» текста, активно разрабатываемые в научной литературе [4, 10, 18]

и построенные как на использовании словарей, так и на алгоритмах классификации на основе обучения.

Эмоциональная составляющая, выраженная на уровне лексемы или коммуникативного фрагмента, называется лексической тональностью. Тональность текста в целом определяется лексической тональностью составляющих его единиц и правилами их сочетания.

Анализ тональности текста в [18] реализуется в несколько этапов: все слова (существительные, глаголы, прилагательные и наречия) и некоторые словосочетания размечаются по заранее подготовленным словарным спискам тональной лексики. Каждому слову приписывается сила тональности. Если слово не нашлось в списках тональной лексики, то оно считается нейтральным. После этого на основе неполного синтаксического анализа слова и словосочетания объединяются в тональные цепочки. Тональные словари заполняются экспертами и содержат более 15 000 тональных слов и коллокаций. Слова и словосочетания соединяются по правилам сочетаемости лексем между собой. В результате анализа предложения в тексте приводятся к одному из типов синтаксической структуры из субъекта, предиката и объекта, где каждый член структуры в общем случае представлен цепочкой словоформ с определенной тональностью. Авторы [18] показывают высокое качество своих методик на текстах определенной тематики, но одновременно указывают на проблемы, связанные с ошибками, в том числе и лексического анализа.

Метод автоматического определения тональности, предложенный в [21], построен на использовании семантических шаблонов, которые являются ориентированными графами, представляющими собой фрагменты деревьев синтаксической зависимости с ограничениями, наложенными на его вершины. Существенно, что ограничения, накладываемые на узлы семантических шаблонов, дополнены специальными словарями (фильтрами), содержащими лексику, выражающую позитивную или негативную оценку. Очевидно, что положительные результаты тестирования связаны во многом с узкоспециальным набором текстовых сообщений, к которым данная методика применяется.

Отметим, что большинство работ, в которых рассматривается задача определения «тональности» текста [6, 18, 21, 23, 24], решают примитивную задачу разделения на два, или три класса: положительные, нейтральные и отрицательные. В такой постановке разработанные методы не применимы для решения сложных задач интеллектуального анализа данных, таких как определение различных типов экстремистской направленности текстов.

Из рассмотренных работ видно, что необходимым условием для анализа тональности является составление словарного списка тональной лексики. Методика классификации может опираться на сложную словарную систему, как в [4, 10]. В работе [10] метод автоматической классификации основан на словаре эмоциональной лексики, в работе [4] авторы надеются на словарь и семантический классификатор В. А. Тузова. По сути, выбор и формирование такого рода словарей представляет собой определение дифференцирующих признаков для конкретных методик.

Методы интеллектуального анализа текста

Для перехода в задаче анализа текстов от использования простейших лексических единиц (как описано выше) к более сложным признакам предполагает использование реляционно-ситуационной модели текста [17], представляющей семантику текста в виде множества объектов с семантическими ролями и связями между ними. Реляционно-ситуационная модель текста строится на основе синтактико-семантического анализа текстов. Реляционно-ситуационный анализ состоит в переходе от синтаксем к их значениям и от значения синтаксем к значениям высказываний. Установление значений синтаксем позволяет формировать наборы сложных психолингвистических и семантических признаков [1], использование которых для рассматриваемых здесь задач анализа русскоязычных текстов требует дополнительных объемных исследований.

В работе [29] разрабатывались методы автоматический классификатор сюжетов на основе поверхностного синтаксического анализа (определения частей речи и именованных существей), а также семантических признаков. Сюжетом считается фрагмент текста, в котором описывается некоторое действие, его исполнитель и результат. Алгоритм обучался на корпу-

се из 16 930 текстов собранных с сайтов исламистских экстремистов. Тексты отбирались экспертами и размечались вручную на “сюжетные” и “несюжетные”.

Для выявления скрытых структурных инвариантов текстов для задач выявления экстремизма в [25] рассматривается психосемантический подход. Психосемантика [19] рассматривает семантическое пространство понятий, которое может представлять собой совокупность лексических единиц вместе с их ассоциациями. Вербальная семантика в [19] изучает метод семантического дифференциала, предложенный американскими психологами во главе с Чарльзом Осгудом. Он относится к методам шкалирования и получил широкое распространение в психолингвистике и экспериментальной психосемантике. Методы шкалирования используются в психологии с целью получения количественных показателей для оценки отношения испытуемых к определенным объектам. Семантический дифференциал в психолингвистике – это метод количественного и качественного индексирования значения слова с помощью двухполюсных шкал, в которой имеется градация с помощью антонимичных прилагательных. Техника построения и анализа семантических пространств включает применение математических процедур факторного и кластерного анализа, многомерной статистики, структурного моделирования.

В рамках разработанной В. Ф. Петренко парадигмы [19] был проведен ряд масштабных исследований семантических пространств различных политических партий, картин мира различных этнических, социальных, профессиональных сообществ, религиозных конфессий, особенностей восприятия произведений искусства (живописи, художественных фильмов), специфики измененных состояний сознания.

В [19] была заявлена программа исследований методами экспериментальной психосемантики религиозной картины мира в контексте борьбы с терроризмом. Предлагалось базовое семантическое пространство строить на основе оценок текстовых фрагментов религиозных канонических текстов носителями культа и националистами. На основе оценок можно построить семантическое пространство, которое будет отражать ценностные установки и семантические пространства для религиозных фанатиков и носителей тоталитарных идеологий. Логично предположить возможность использования построенных семантических пространств для анализа различных наборов текстов.

Дифференцирующие признаки тематики текста

Выбор дифференцирующих признаков для методов классификации и рубрикации текстов является ключевой проблемой для создания методик выявления тематической направленности текстов на естественных языках сети Интернет [3, 13, 20, 28]. Для текстов на естественных языках в качестве признаков обычно рассматриваются слова и взаимосвязанные наборы слов, содержащиеся в текстах. Для задачи выявления текстов экстремистской направленности существенно определение набора анализируемых признаков, таких, например, как результаты полного лингвистического анализа, ключевая лексика, выделенные именованные сущности [1, 14, 29]

В [5] исследовались возможности создания методов автоматического выявления агрессивности в текстах социальных сетей. Выполнено выявление психолингвистических характеристик текста и определение процентного содержания слов и словосочетаний из заданных словарей.

В [14, 20, 28] проводились эксперименты на реальных данных для текстов на русском, английском, татарском и башкирском языках по выявлению текстов по тематикам: «наркотики», «насилие», «национализм», «отрицание ценностей», «порнография», «терроризм», «фашизм», «экстремизм». Были выявлены наборы дифференцирующих признаков, включающих такие признаки, как существительные, именные группы, прилагательные. Анализ результатов тестирования показал, что можно использовать псевдоосновы, выделенные аналитическим алгоритмом морфологического анализа, для автоматического рубрицирования и классификации текстовых сообщений по специальным тематикам, связанным с нарушениями предоставления информации в Интернете.

В [1] экспериментально решалась задача выявления лексических, психолингвистических и семантических признаков текстов, позволяющих с помощью методов машинного обучения

выявлять материалы экстремистского содержания. Эмпирическим путем было показано, что представленные психолингвистические и семантические характеристики (маркеры) являются значимыми признаками для решения задачи выявления экстремистских текстов.

В работах К. В. Злоказова [8, 9] изучаются тексты деструктивной направленности путем выделения признаков объекта, цели и процесса деструкции, выявления эмоционально акцентированных слов и отношения автора к объекту и цели деструкции. Поиск различий по параметрам лексических, грамматических и морфологических характеристик текстов, проведенный с помощью однофакторного дисперсионного анализа, результатов не дал. В [8] указывается на несколько причин: данные показатели свидетельствуют о сопоставимости когнитивных уровней авторов, категория деструкции воплощена в семантике текста и должна соотноситься с элементами подобной природы. Делается вывод о том, что необходимо описание семантики деструкции в тексте с использованием категорий объекта, цели и процесса.

Попытка оценки различных грамматических форм в качестве дифференцирующих признаков для тематической классификации текстов экстремистской направленности на русском языке проводилась нами методом Relief [22, 32], который применяется для отбора признаков, используемых в задаче бинарной классификации. Классифицируемые объекты представлены p признаками. Пусть S – обучающая выборка объектов размера n . F – заданный набор признаков (f_1, \dots, f_p) . Объект X задается p -мерным вектором (x_1, \dots, x_p) , где x_j – значение признака f_j . Пусть для каждого из признаков определена функция $diff(x_k, y_k)$, принимающая значения на множестве $[0, 1]$, где значение 1 обозначает совпадение значений признаков, а 0 – несовпадение. Метод итеративно аппроксимирует значимость (r_1, \dots, r_p) признаков. Значение r_p лежит в интервале $[0, 1]$, где 0 обозначает отсутствие значимости признака, а 1 – высокую значимость. На каждой итерации метод выбирает случайным образом объект X , а также ближайший к нему (по метрике L1) пример $Y^- = (y_1^-, \dots, y_p^-)$, из отрицательного и $Y^+ = (y_1^+, \dots, y_p^+)$ положительного класса. Вектор $R = (r_1, \dots, r_p)$ значимости признаков обновляется следующим образом:

$$r_i \leftarrow r_i - \|diff(x_i, y_i^+)\|_1 + \|diff(x_i, y_i^-)\|_1.$$

После m итераций (m – параметр запуска метода) значимость нормируется в отрезок $[0, 1]$. Таким образом, незначимые признаки получают значение близкое к 0, а значимые – существенно отличное от 0.

Метод Relief был применен для оценки значимости частотных признаков, отвечающих именным и глагольным группам из текстов обучающей выборки, описанной в [1]. Значение каждого признака – относительная частота встречаемости в тексте. К сожалению, применение метода в нашем случае дало сопоставимо высокое значение значимости для почти всех признаков, что не позволило выделить подмножество значимых признаков небольшого размера.

Заключение

Рассмотренные публикации позволяют сформулировать следующие проблемы, возникающие при решении задачи выявления экстремистской направленности в текстах на русском языке.

- Психологами и лингвистами не разработаны применимые для автоматического анализа текстов критерии экстремистской направленности текстов.
- Разработанные методика опираются в основном на словарные системы, что ограничивает их применение в реальных информационных системах, требующих огромных затрат на актуализацию словарей, баз данных и большого количества правил.
- В основном используется бинарная классификация (или разделение на три класса), что не решает в полной мере задачу выявления сложной тематической и эмоциональной направленности текстов на естественных языках.

- Недостаточно исследованы дифференцирующие признаки, определяющие экстремистскую направленность текстов при их автоматической классификации и рубрикации.

Автоматическое выделение текстов противоправной тематики требует серьезных исследований, прикладных и фундаментальных разработок.

Список литературы

1. Ананьева М. И., Девяткин Д. А., Кобозева М. В., Смирнов И. В. Лингвостатистический анализ текстов экстремистской направленности // Тр. Междунар. науч. конф. Московского физико-технического института (государственного университета) и Института физико-технической информатики (SCVRT1516). М.; Протвино, 2016. С. 210–213.
2. Бакина А. В., Махова И. Ю. К вопросу о психологических критериях экспертной оценки экстремистской направленности текста // Наука и мир. 2014. № 2. С. 178–183.
3. Батура Т. В. Формальные методы определения авторства текста // Вестн. НГУ. Серия: Информационные технологии. 2012. Т. 10, вып. 4. С. 81–94.
4. Боярский К. К., Каневский Е. А., Саганенко Г. И. К вопросу автоматической классификации текстов // Экономико-математические исследования: математические модели и информационные технологии. VII. СПб.: Нестор-История, 2009. С. 252–273.
5. Девяткин Д. А., Кузнецова Ю. М., Чудова Н. В., Швец А. В. Интеллектуальный анализ проявлений вербальной агрессивности в текстах сетевых сообществ // Искусственный интеллект и принятие решений. 2014. № 2. С. 27–41.
6. Жданова С. Ю., Мишланова С. Л., Поляков В. Б., Рабчевский Е. А. Особенности репрезентации этнической агрессии в корпусе сообщений пермского сегмента социальной сети «ВКонтакте» (VK.COM) // Вектор науки Тольяттинского государственного университета. Серия: Педагогика, психология. 2012. № 4 (11). С. 246–248.
7. Зеленина О. В., Сулонов П. Е. Методика выявления признаков экстремизма. Процессуальные исследования (экспертизы) аудио-, видео- и печатных материалов: Науч.-практ. пособие. Екатеринбург, 2009. 90 с.
8. Злоказов К. В. Контент-анализ текстов деструктивной направленности // Политическая лингвистика. 2015. № 1. С. 244–251.
9. Злоказов К. В. Восприятие экстремистского текста субъектами с различным уровнем деструктивной установки // Политическая лингвистика. 2014. № 1. С. 265–272.
10. Клековкина М. В., Котельников Е. В. Метод автоматической классификации текстов по тональности, основанный на словаре эмоциональной лексики // Тр. 14-й Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL-2012. Переславль-Залесский, 2012. С. 81–86.
11. Кобзева С. В. Противодействие распространению агрессивной информации: Мировой опыт. М.: Либроком, 2009. 64 с.
12. Маннинг К., Рагхаван П., Шютце Х. Введение в информационный поиск. М.: ООО «И.Д.Вильямс», 2014. 528 с.
13. Мбайкоджи Э., Драль А. А., Соченков И. В. Метод автоматической классификации коротких текстовых сообщений // Информационные технологии и вычислительные системы. 2012. № 3. С. 93–102.
14. Михайлов А. С., Соколова Т. В., Чеповский А. А., Чеповский А. М. Выявление тематической направленности текстов на естественных языках // Искусственный интеллект и принятие решений. 2016. № 1. С. 9–17.
15. Мишланов В. А. Законодательство РФ об экстремизме и задачи лингвистической экспертизы текстов // Вестн. Перм. ун-та. Российская и зарубежная филология. 2012. № 3 (19). С. 62–69.
16. Мишланов В. А., Салимовский В. А. Этнический экстремизм в массовой коммуникации с точки зрения проблем судебной лингвистической экспертизы // Вестн. Перм. ун-та. Российская и зарубежная филология. 2013. Вып. 4 (24). С. 63–75.
17. Осипов Г. С. Методы искусственного интеллекта. М.: Физматлит, 2011. 296 с.
18. Пазельская А. Г., Соловьев А. Н. Метод определения эмоций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегод-

ной Международной конференции «Диалог» (Бекасово, 25–29 мая 2011 г.). М.: Изд-во РГГУ, 2011. Вып. 10 (17). С. 574–586.

19. *Петренко В. Ф.* Основы психосемантики. 3-е изд. М.: Эксмо, 2010. 480 с.

20. *Поляков И. В., Соколова Т. В., Чеповский А. А., Чеповский А. М.* Проблема классификации текстов и дифференцирующие признаки // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2015. Т. 13, № 2. С. 55–63.

21. *Поляков П. Ю., Калинина М. В., Плешко В. В.* Опыт построения системы автоматического определения тональности объектов на основе синтактико-семантического анализатора // Электронные библиотеки. 2015. Т. 18, № 3–4. С. 184–202.

22. Реализация алгоритма ReliefF. URL: <https://github.com/rhiever/ReliefF> (дата обращения 20.06.2016).

23. *Рубцова Ю. В.* Построение корпуса текстов для настройки тонового классификатора // Программные продукты и системы. 2015. № 1 (109). С. 72–78.

24. *Салимовский В. А., Ермакова Л. М.* Экстремистский дискурс в массовой коммуникации Рунета // Вестн. Перм. ун-та. Российская и зарубежная филология. 2011. № 3 (15). С. 71–80.

25. Скрытое эмоциональное содержание текстов СМИ и методы его объективной диагностики / Под ред. А. А. Леонтьева, Д. А. Леонтьева. М.: Смысл, 2004. 229 с.

26. *Фаткуллин Б. Г.* Использование теории множеств в сравнительно-логических методах выделения текстов на исламскую тематику в процессе мониторинга сетевых ресурсов // Вестн. ЮУрГУ. Серия «Лингвистика». 2016. Т. 13, № 3. С. 22–26.

27. *Холопова Е. Н., Аринцева И. К.* Выявление материалов с признаками экстремизма // Вестн. Калининград. филиала АНОВО «Международный университет в Москве»: Сб. ст. / Под ред. А. А. Ярцева. Калининград: Искра, 2015. Вып. 6. С. 185–191.

28. *Чеповский А. М.* Информационные модели в задачах обработки текстов на естественных языках. 2-е изд., перераб. М., 2015. 276 с.

29. *Ceran B. et al.* A semantic triplet based story classifier // Proc. of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012). IEEE Computer Society, 2012. P. 573–580.

30. *Chen H.* Exploring extremism and terrorism on the web: the dark web project // Pacific-Asia Workshop on Intelligence and Security Informatics. Springer Berlin Heidelberg, 2007. С. 1–20.

31. *Finlayson M. A., Halverson J. R., Corman S. R.* The N2 corpus: A semantically annotated collection of Islamist extremist stories // LREC. 2014. С. 896–902.

32. *Kira K., Rendell L. A.* The feature selection problem: Traditional methods and a new algorithm // AAAI. 1992. Т. 2. С. 129–134.

Материал поступил в редколлегию 25.10.2016

**M. I. Ananyeva¹, M. V. Kobozeva¹, F.N Solovyev²
I. V. Polyakov³, A. M. Chepovskiy³**

¹ *ISA FRC Computer Science & Control RAS
9 Prospekt 60-letiya Oktyabrya, Moscow, 117312, Russian Federation*

² *Moscow Polytechnic University
38 Bolshaya Semenovskaya Str., Moscow, 107023, Russian Federation*

³ *National Research University Higher School of Economics
20 Myasnitskaya Str., Moscow, 101000, Russian Federation*

achepovskiy@hse.ru

THE PROBLEM OF DETECTION OF EXTREMIST TEXTS

This paper reviews the methods used for detection of extremist texts from Internet. Further research is needed to find a new methods of text classification and features for extremist texts classification.

Keywords: text classification, extremist texts.

References

1. Ananyeva M.I., Devyatkin D.A., Kobozeva M.V., Smirnov I.V. Lingvostatisticheskij analiz tekstov ehkstremitstskoj napravlenosti // V kn.: Trudy Mezhdunarodnoj nauchnoj konferencii Moskovskogo fiziko-tehnicheskogo instituta (gosudarstvennogo universiteta) i Instituta fiziko-tehnicheskoy informatiki (SCVRT1516). M., Protvino: Institut fiziko-tehnicheskoy informatiki, 2016. – S. 210 – 213.
2. Bakina A.V., Mahova I.YU. K voprosu o psihologicheskikh kriteriyah ehkspertnoj ocenki ehkstremitstskoj napravlenosti teksta // Nauka i mir, 2014. – №2. – S. 178 – 183.
3. Batura T.V. Formal'nye metody opredeleniya avtorstva teksta // Vestnik NGU. Seriya: Informacionnye tekhnologii. – 2012. – Tom.10. – Vypusk 4. – S. 81 – 94.
4. Boyarskij K.K., Kanevskij E.A., Saganenko G.I. K voprosu avtomaticheskoy klassifikacii tekstov / V kn.: Ehkonomiko-matematicheskie issledovaniya: matematicheskie modeli i informacionnye tekhnologii. VII. – SPb: SPb EHMI RAN. Nestor-Istoriya, 2009. – S. 252 – 273.
5. Devyatkin D.A., Kuznecova YU.M., Chudova N.V., Shvec A.V. Intellektual'nyj analiz proyavlenij verbal'noj agressivnosti v tekstah setevyh soobshchestv // Iskusstvennyj intellekt i prinyatie reshenij. 2014. – № 2. – S. 27-41.
6. Zhdanova S.YU., Mishlanova S.L., Polyakov V.B., Rabchevskij E.A. Osobennosti reprezentacii ehlnicheskoy agressii v korpuse soobshchenij permskogo segmenta social'noj seti «Vkontakte» (VK.COM) // Vektor nauki Tol'yattinskogo gosudarstvennogo universiteta. Seriya: Pedagogika, psihologiya. – 2012. – №. 4 (11). – S. 246 – 248.
7. Zelenina O.V., Suslonov P.E. Metodika vyyavleniya priznakov ehkstremitzma. Processual'nye issledovaniya (ehkspertizy) audio-, video – i pechatnyh materialov. Nauchno-prakticheskoe posobie. – Ekaterinburg: Ural'skij yuridicheskij institut MVD Rossii, 2009. – 90 s.
8. Zlokazov K.V. Kontent-analiz tekstov destruktivnoj napravlenosti // Politicheskaya lingvistika. – 2015. – №. 1. – S. 244 – 251.
9. Zlokazov K.V. Vospriyatie ehkstremitstskogo teksta sub"ektami s razlichnym urovnem destruktivnoj ustanovki // Politicheskaya lingvistika. – 2014a. – №1. – S. 265 – 272.
10. Klekovkina M.V., Kotel'nikov E.V. Metod avtomaticheskoy klassifikacii tekstov po tonal'nosti, osnovannyj na slovare ehmocional'noj leksiki / Trudy 14-j Vserossijskoj nauchnoj konferencii «EHlektronnye biblioteki: perspektivnye metody i tekhnologii, ehlektronnye kollekicii» – RCDL-2012, Pereslavl'-Zalesskij, Rossiya, 15-18 oktyabrya 2012 g. – C. 81–86.
11. Kobzeva S.V. Protivodejstvie rasprostraneniya agressivnoj informacii: Mirovoj opyt. M.: Knizhnyj dom «LIBROKOM», 2009. – 64 s.
12. Manning C., Raghavan P., Shutce H. Vvedenie v infor-macionnyj poisk. M.: OOO «I.D.Vil'yams», 2014. – 528 s.
13. Mbajkodzhi EH., Dral' A.A., Sochenkov I.V. Metod avtomaticheskoy klassifikacii korotkih tekstovyh soobshchenij. // Informacionnye tekhnologii i vychislitel'nye sistemy (M.). – 2012. – №3. – S. 93 – 102.
14. Mihaylov A.S., Sokolova T.V., Chepovskiy A.A., Chepovskiy A.M. Vyyavlenie tematicheskoy napravlenosti tekstov na estestvennyh yazykah // Iskusstvennyj intellekt i prinyatie reshenij. 2016. – № 1. – S. 9 – 17.
15. Mishlanov V.A. Zakonodatel'stvo RF ob ehkstremitzme i zadachi lingvisticheskoy ehkspertizy tekstov // Vestnik Permskogo universiteta. Rossijskaya i zarubezhnaya filologiya. – 2012. – №. 3(19). – S. 62 – 69.
16. Mishlanov V.A., Salimovskij V.A. EHlnicheskij ehkstremitizm v massovoj kommunikacii s tochki zreniya problem sudebnoj lingvisticheskoy ehkspertizy // Vestn. Perm. un-ta. Rossijskaya i zarubezhnaya filologiya. – 2013. – Vyp. 4 (24). – S. 63 – 75.
17. Osipov G.S. Metody iskusstvennogo intellekta. – M.: FIZMATLIT. – 2011. – 296 s.
18. Pazel'skaya A.G., Solovyev A.N. Metod opredeleniya ehmocii v tekstah na russkom yazyke / Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Po materialam ezhegodnoj Mezhdunarodnoj konferencii «Dialog» (Bekasovo, 25 – 29 maya 2011 g.). Vyp. 10 (17). – M.: Izdvo RGGU, 2011. – S. 574 – 586.

19. Petrenko V.F. *Osnovy psihosemantiki*. 3-e izdanie. – M.: Ehksimo, 2010. – 480 s.
20. Polyakov I.V., Sokolova T.V., Chepovskiy A.A., Chepovskiy A.M. Problema klassifikacii tekstov i differenciruyushchie priznaki // *Vestnik Novosibirskogo gosudarstvennogo universiteta*. Seriya: Informacionnye tekhnologii. – 2015. – T. 13. – № 2. – S. 55 – 63.s.
21. Polyakov P.Yu., Kalinina M.V., Pleshko V.V. Opyt postroeniya sistemy avtomaticheskogo opredeleniya tonal'nosti ob'ektov na osnove sintaktiko-semanticheskogo analizatora // *Elektronnye biblioteki*. – 2015. – T. 18. – № 3-4. – S. 184 – 202.
22. Realizatsiya algoritma Relief. URL: <https://github.com/rhiever/Relief> (data obrashcheniya: 20.06.2016).
23. Rubcova Yu.V. Postroenie korpusa tekstov dlya nastroyki tonovogo klassifikatora // *Programmnye produkty i sistemy*. – 2015. – №1(109). – S.72 – 78.
24. Salimovskij V.A., Ermakova L.M. Ekhstremistskij diskurs v massovoj kommunikacii Runeta // *Vestnik Permskogo universiteta*. Rossijskaya i zarubezhnaya filologiya. – 2011. – №. 3(15). – S. 71 –80.
25. Skrytoe ehmocional'noe sodержanie tekstov SMI i metody ego ob'ektivnoj diagnostiki / Pod red. A.A. Leont'eva, D.A. Leontyeva. – M.: Smysl, 2004. – 229 s.
26. Fatkulin B.G. Ispol'zovanie teorii mnozhestv v sravnitel'no-logicheskikh metodah vydeleniya tekstov na islamskuyu tematiku v processe monitoringa setevyh resursov // *Vestnik YUUrGU*. Seriya «Lingvistika». – 2016. – T. 13. – №. 3. – S. 22 – 26.
27. Holopova E.N., Arinceva I.K. Vyyavlenie materialov s priznakami ehkstremizma // *Vestnik Kaliningradskogo filiala ANOVO «Mezhdunarodnyj universitet v Moskve»*: sb. st. / Pod red. A.A. Yarceva. – Kaliningrad: Izd-vo «Iskra». – 2015. – Vyp. 6. – С. 185 – 191.
28. Chepovskiy A. M. Informacionnye modeli v zadachah obrabotki tekstov na estestvennyh yazykah. Vtoroe izdanie, pererabotannoe. M.: Nacional'nyj otkrytyj universitet «INTUIT», 2015. – 276 s.
29. Ceran B. et al. A semantic triplet based story classifier // *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*. – IEEE Computer Society, 2012. – C. 573 – 580.
30. Chen H. Exploring extremism and terrorism on the web: the dark web project // *Pacific-Asia Workshop on Intelligence and Security Informatics*. – Springer Berlin Heidelberg, 2007. – C. 1 – 20.
31. Finlayson M.A., Halverson J.R., Corman S.R. The N2 corpus: A semantically annotated collection of Islamist extremist stories // *LREC*. – 2014. – C. 896 – 902.
32. Kira K., Rendell L.A. The feature selection problem: Traditional methods and a new algorithm // *AAAI*. – 1992. – T. 2. – C. 129 – 134.