

А.Х. Хакимова¹
 М.М. Шарнин²
 С.В. Клименко³
 О.В. Золотарев⁴
 И.В. Родина⁵

A.Kh. Khakimova
 M.M. Charnine
 S.V. Klimenko
 O.V. Zolotarev
 I.V. Rodina

**МЕРА ПОДОБИЯ ТЕКСТОВ
 КАК ИНСТРУМЕНТ ОЦЕНКИ
 ИНТЕРТЕКСТУАЛЬНОСТИ ПРИ АНАЛИЗЕ
 БОЛЬШИХ КОЛЛЕКЦИЙ ДОКУМЕНТОВ⁶**

**THE MEASURE OF TEXT SIMILARITY
 AS A TOOL FOR ASSESSMENT
 OF INTERTEXTUALITY IN LARGE
 COLLECTIONS OF DOCUMENTS**

Предложена методика обнаружения интертекстуальных отношений через нахождение неявных ссылок с помощью лингвистических и статистических методов. Интертекстуальность – это наличие в тексте элементов и идей из других текстов. Продемонстрирована возможность выявления временной межъязыковой миграции терминов и идей с целью прогноза и определения идеологических траекторий. Предложена новая мера подобия текстов, апробированная на коллекции научных документов и развиваемая путем максимизации корреляции явных и неявных ссылок. Описана методика кластеризации документов в соответствии с мерой подобия текстов. Продемонстрирована возможность применения предложенной меры для анализа экстремистских текстов из Интернета.

Ключевые слова: интертекстуальность, миграция идей, мера подобия, кластеризация документов, неявные ссылки.

The method for detection of intertextual relations by finding implicit links using linguistic and statistical methods is suggested. Intertextuality is a presence in one text of elements and ideas from other texts. The possibility of identifying of cross-language migration of terms and ideas for prognosis and determination of ideological trajectories is demonstrated. A new text similarity measure is suggested. The measure was tested using collection of scientific documents. The measure was improved by maximizing correlation between explicit and implicit links. A method for documents clustering according to the measure of text similarity is suggested. The possible application of the proposed measure for analysis of extremist texts from the Internet is suggested.

Keywords: intertextuality, migration of ideas, measure of similarity, texts clustering, implicit links.

1. Скрытая интертекстуальность

В литературе по анализу текстов интертекстуальность определяется как наличие явного или скрытого присутствия одного текста внутри другого [15]. Интертекстуальность создает взаимосвязь между текстами и генерирует соответствующее понимание читателем, опираясь на

¹ Кандидат биологических наук, ведущий научный сотрудник АНО ИФТИ.

² Кандидат технических наук, ведущий научный сотрудник Института проблем информатики Российской академии наук Федерального исследовательского центра «Информатика и управление» Российской академии наук.

³ Доктор физико-математических наук, профессор, генеральный директор АНО ИФТИ.

предшествующий уровень знаний и понимания читателей. Примером интертекстуальности являются заимствования, трансформация предшествующего текста без использования явных ссылок.

Явная межсмысловая интертекстуальность играет законную роль в научных публикациях

⁴ Кандидат технических наук, доцент, заведующий кафедрой информационных систем в экономике и управлении АНО ВО «Российский новый университет».

⁵ Ведущий научный сотрудник АНО ИФТИ.

⁶ Работа выполнена при поддержке Российского фонда фундаментальных исследований, гранты 16-07-00756 и 16-29-09527. Мы благодарны РФФИ за поддержку и финансирование наших проектов.

(цитаты из оригинальных публикаций на ту же тему, из связанных с темой публикаций, ссылки). Тем не менее, скрытая интертекстуальность является извечной проблемой (Bouville 2008).

Мы представляем новые методики, способные обнаружить скрытую интертекстуальность в научной литературе, а также выявить её наличие в сходных текстах и измерить степень выраженности. Возможность обнаружения интертекстуальных отношений через нахождение неявных ссылок позволит обнаружить смысловое подобие текстов.

Неявными ссылками мы считаем дублирование значительной части одного или нескольких оригинальных текстов без использования ссылки на оригиналы. На практике встречаются случаи импортирования оригинального текста в последующий текст с незначительными изменениями, а также более скрытые случаи, например изложение оригинальной идеи другими словами.

Проблема интертекстуальности в научных публикациях нуждается в серьезном изучении. Когда ученый обращается к новой теме, первым шагом является поиск оригинальных публикаций в этой области. Установление происхождения идей является важным условием для развития знаний и обмена концепциями, инструментов и данных между исследователями. Скрытая интертекстуальность осложняет исследование и подрывает доверие между исследователями [15].

Анализ интертекстуальности в больших коллекциях документов помогает решить следующие задачи: поиск аналогичных документов, определение перспективных направлений и научных школ, выявление наиболее влиятельных документов и потоков идей в заданной предметной области, в частности – помогает решить такую актуальную задачу, как мониторинг экстремистской и террористической идеологии в Интернете.

2. Интернет как среда распространения террористической информации

Интернет обладает рядом особенностей, которые подталкивают к его многогранному использованию, включая:

- легкость доступа;
- незначительные масштабы госрегулирования и цензуры или их отсутствие;
- наличие огромной всемирной потенциальной аудитории пользователей;
- анонимность связи;
- высокая скорость передачи информации;
- мультимедийность среды [29].

Терроризм в Интернете распространяется далеко за пределы веб-сайтов, непосредственно

управляемых или контролируемых террористическими организациями. Их сторонники используют все доступные инструменты в Интернете (новостные агентства, чаты, блоги, форумы, группы пользователей, социальные сети и виртуальные миры) [8]. Прекращение деятельности таких интернет-ресурсов зачастую невозможно в силу правовых и юридических сложностей, а иногда малоэффективно, так как многие сайты специально постоянно меняют свой адрес. Число подобных страниц постоянно растет, что может косвенно свидетельствовать об увеличении количества сторонников радикальных группировок [32].

Социальные сети являются реальным информационным плацдармом, на котором развивается политическая и социальная борьба. Политологи используют официальный термин «твиттер-революция» [35], признавая, что сервис микроблогов Твиттер – это реальный инструмент, который эффективно используют протестующие в различных государствах. Именно Twitter стал стартовой площадкой для цифрового строительства всемирного халифата [34].

Управление Организации Объединенных Наций по наркотикам и преступности (UNODC) представило классификацию средств, с помощью которых Интернет используется для содействия и поддержки террористических актов. Были идентифицированы шесть иногда пересекающихся категорий:

- пропаганда террористической деятельности;
- финансирование деятельности;
- обучение членов террористических группировок;
- планирование террористических актов;
- координация и исполнение террористических актов;
- кибератаки [20].

Одно из основных применений Интернета террористами – это распространение пропаганды. Информация может подаваться в форме исторической справки, сведений о лидерах, манифестов и т.п. Но террористы могут использовать Интернет также для распространения дезинформации, передачи угроз.

Как утверждали Weimann и Winn, одной из центральных проблем, стоящих перед терроризмом, является оправдание применения насилия [23]. Террористическая риторика пытается переложить ответственность на противника, показывая его жестокость, бесчеловечность, безнравственность. Многие сайты утверждают, что их конечной целью является дипломатическое уре-

гулирование, достигнутое путем переговоров и международного давления на репрессивное правительство.

Использование поисковых систем Интернета позволяет легко идентифицировать и получить связанное с терроризмом содержание. В последние годы террористические организации всё чаще используют Интернет как альтернативный учебный полигон [6; 13; 18; 20]. Сети позволяют создавать новую форму террористических организаций, действующих в децентрализованной манере, в отличие от традиционного иерархического построения. Внутри сетей нет иерархии, имеется множество лидеров в зависимости от размеров группы [28].

В последние годы в качестве отличительного признака терроризма всё чаще упоминается не только политическая, но и идеологическая, в том числе и религиозная, мотивация насилия [25].

В настоящее время террористами для создания идеологической платформы своей деятельности используется преимущественно идеология так называемого радикального ислама. Но терроризм использует многие другие неофашистские, национал-сепаратистские, религиозные течения.

Многие экстремистские и террористические сайты поддерживаются на нескольких языках. В последние годы экстремистская пропаганда всё чаще ведется на русском языке, который является универсальным языком общения на территории бывшего СССР.

Интернет сегодня превратился в мощный инструмент манипуляции сознанием и поведением представителей молодежной среды в силу более низкого порога критического восприятия всего нового, максималистского восприятия окружающей социальной реальности и стремления к протестной идентификации.

Как указывают ученые-правоведы, внедрение в ментальность россиян жестокости, идеи о допустимости применения насилия создают благоприятную почву для распространения в стране экстремизма и терроризма [26]. Как отметил первый заместитель руководителя аппарата НАК, к.ю.н. Ильин Е.П.: «Вместо трудоемкого и долгого изучения теологических наук, позволяющих сформировать целостную картину мира и религии, идеологи терроризма предлагают набор вырванных из контекста цитат Корана и хадисов, обосновывающих необходимость постоянной борьбы с «кафирами» и «мунафиками» [27]. Русский язык занимает в объеме пропаганды ИГИЛ/ДАИШ третье место после арабского и английского.

По результатам исследований ученых ПФУ и Центра исламоведческих исследований АН РТ, были определены целевые аудитории пропагандистов экстремистской идеологии. Так, молодые мусульмане в возрасте до 30 лет в качестве основного источника знаний по исламу указали Интернет. Мониторинг ряда исламских интернет-сообществ в русскоязычном сегменте сети Интернет показал, что число участников данных тематических групп, в которых осуществляются призывы к построению халифата, постоянно растет: рост аудитории из числа жителей Российской Федерации за период с января по сентябрь 2015 года составил порядка 56 000 человек – со 120 907 до 176 961. При этом 69% от общего числа участников групп составляют люди в возрасте до 30 лет [24].

Несмотря на постоянные смены адреса экстремистских сайтов, существуют методы их автоматической идентификации. Целью данной работы является разработка методик, позволяющей автоматически идентифицировать экстремистские сайты, их взаимосвязь и влияние (т.е. выявлять наиболее влиятельные экстремистские сайты) для мониторинга, прогнозирования и визуализации потоков экстремистских идей. Известно большое количество научных работ, анализирующих влияние научных идей и потоки научных идей, которые основываются на анализе явных и неявных ссылок между статьями [7; 10–12; 17]. Аналогичные подходы на основе неявных ссылок применимы также к анализу экстремистских сайтов, особенностью которых является почти полное отсутствие явных, т.е. формальных библиографических и гипертекстовых ссылок на источники идей.

3. Организация противодействия терроризму в Интернете в России

Основные базовые понятия, раскрывающие суть системы противодействия идеологии терроризма в РФ воплощены в Федеральном законе № 35-ФЗ «О противодействии терроризму» и Указе Президента Российской Федерации «О мерах по противодействию терроризму».

В Федеральном законе «О противодействии терроризму» дано определение понятию «терроризм», включившее в свой состав термин «идеология насилия» [1]. Указанная формулировка послужила законодательным обоснованием необходимости существования наряду с институтами борьбы с терроризмом институтов предупреждения и профилактики терроризма.

Террористические группировки используют новейшие информационные технологии и Интернет для распространения пропаганды и обме-

на информацией, привлечения финансовых ресурсов, планирования террористических актов, а также для осуществления контроля за их проведением. Этот фактор определен Концепцией противодействия терроризму в Российской Федерации, утвержденной Президентом РФ 5 октября 2009 г., среди основных, способствующих возникновению и распространению терроризма [5].

Согласно Концепции, одним из направлений профилактики терроризма является противодействие распространению идеологии терроризма путем обеспечения защиты единого информационного пространства Российской Федерации; совершенствование системы информационного противодействия терроризму.

Вместе с тем, необходимо отметить, что единственное, на данный момент, определение идеологии терроризма дано в Комплексном плане противодействия идеологии терроризма в Российской Федерации на 2013–2018 годы, утвержденном Президентом Российской Федерации 26 апреля 2013 года (№ Пр-1069) [5].

Под идеологией терроризма (террористическая идеология) понимаются совокупность идей, концепций, верований, догматов, целевых установок, лозунгов, обосновывающих необходимость террористической деятельности и направленных на мобилизацию людей для участия в этой деятельности.

В качестве одного из направлений противодействия идеологии терроризма рассматривается создание и задействование механизмов защиты информационного пространства от проникновения идей, оправдывающих террористическую деятельность [5].

Противодействие идеологии терроризма в России организовано на федеральном и региональном уровнях государственного управления с задействованием возможностей местного самоуправления. Решением Президента Российской Федерации главным координатором деятельности по противодействию идеологии терроризма в стране определен Национальный антитеррористический комитет – в соответствии с Федеральным законом от 6 марта 2006 года № 35 «О противодействии терроризму» и Указом Президента России от 15 февраля 2006 года № 116 «О мерах по противодействию терроризму» [1; 4].

24 июня 2016 года Государственная дума Российской Федерации приняла пакет анти-террористических законов, который вводит пожизненное заключение за международный терроризм и обязывает операторов связи, а также мессенджеров и соцсети хранить информацию

о фактах разговоров и переписки пользователей и их содержании. За публичные призывы к терроризму в Интернете, а также его публичное оправдание предусматривается штраф от 300 тысяч до 1 миллиона рублей или лишение свободы от пяти до семи лет с лишением права занимать определенные должности до пяти лет [2; 3].

4. Подходы к мониторингу террористической идеологии в Интернете через выявление неявных ссылок

Наблюдение за тысячами веб-сайтов, дискуссионных форумов, чатов, блогов и других открытых источников в Интернете требует обученного персонала с опытом работы на разных языках, имеющего знакомство с множеством культур, систем верований, политических нюансов и организационных структур террористических групп в Интернете. Такого персонала, разумеется, не хватает, и большинство экспертов сходятся во мнении, что страны должны объединить свои ресурсы [33].

С помощью поисковых систем можно найти некоторые сайты, но нельзя попасть в наиболее радикальные сайты, которые требуют регистрации, пароля или премодерации. После нахождения контента необходимо его проанализировать. Существуют методы, которые способны категоризировать содержание социальных сетей [21]. Перечень базовых информационных технологий противодействия терроризму и их краткое описание представлены, например, И. Котенко [31].

Обычно в корпусе документа содержатся явные и неявные ссылки. Традиционные наукометрические методы включают в себя использование контента сходства; эти отношения могут определяться либо непосредственно, либо косвенно, через кластеризацию терминов или путем использования тезауруса.

Типы явных ссылок, которые предположительно можно обнаружить в документе, будут зависеть от категории документа. Например, в научных статьях можно найти явные ссылки следующего вида:

- 1) цитирование и другие ссылки на документы;
- 2) ссылки на источники данных;
- 3) имена авторов, названия организаций и источников финансирования;
- 4) указатель терминов, глоссарий, тезаурус.

Тексты, содержащие террористическую лексику, обычно не содержат явных библиографических ссылок или гиперссылок, что затрудняет отслеживание взаимосвязи или иерархий идей [39]. В этой связи исследование неявных ссылок в рамках интертекстуальности представляется

единственно возможным методом обнаружения преемственности идей.

В качестве неявных ссылок мы рассматриваем упоминания похожих идей и их авторов [38]. Среди документов могут существовать различные типы ассоциаций. Цель данной работы заключается в том, чтобы выявить неявные ссылки, которые могут быть использованы для изучения и организации документов в коллекции, т.е. представить определенную структуру коллекции.

Неявные ссылки между текстами/сайтами могут быть выявлены с помощью лингвистических и статистических методов. Смысловое сходство фраз определяется с помощью грамматических трансформаций, программ перевода, замен синонимов на термины, полученные с помощью тематического анализа (например, методами LDA и PLSA), а также с помощью ассоциативных связей, выявленных по авторской методике построения ассоциативного портрета предметной области (АППО) [36].

Для выявления неявных ссылок необходимо создание базы данных, содержащей текстовые документы террористического содержания. База данных включает также списки тем, словари терминов, значимые словосочетания и ключевые фразы, составляющие в совокупности лексические ресурсы проекта. При этом связи в базе данных между темами, терминами и фразами будут организованы в виде тезауруса идей. В настоящее время в качестве тестовой базы данных создана и апробируется база данных на тему «Компьютерная графика и виртуальное окружение» [37].

Автоматизированный анализ тестовой БД позволяет строить кластеры, содержащие связан-

ные друг с другом документы. Разработанные оригинальные программные средства позволяют провести анализ общих фраз в документах каждого кластера. Был проанализирован лексический состав кластеров и выявлены слова и фразы, которые встречаются в кластере чаще, чем в большой внешней коллекции на постороннюю тему.

5. Графическое моделирование для прогноза и определения терминологических траекторий

Анализ эволюции семантики в различных предметных областях, взаимодействие терминов и изменение их внутренних свойств может быть выполнен с помощью инструмента анализа больших коллекций книг на разных языках Google Books Ngram. Идеологические траектории определяются как семантические распределения, меняющиеся с течением времени [22]. Анализ позволяет понять динамику употребления терминов в заданной предметной области.

Исследования по миграции терминов показали, что предметные области исследований и знаний эволюционируют, и эта «эволюция» может быть обнаружена. Анализ и мониторинг терминологической миграции позволит обнаружить соответствующие семантические сдвиги. Миграция терминов из английских книг на русский язык получила широкое распространение в 1940–1960 годах. Например, термин «искусственный интеллект» стал использоваться в книгах английского языка в 1957 г. и мигрировал в русскую научную литературу в 1965 г. [9].

Нами была предпринята попытка проследить временную миграцию терминов, несущих идеологическую окраску, на примере термина «кяфир» («кафир»). Идеологи терроризма призывают бороться с неверными (кафирами) [24].

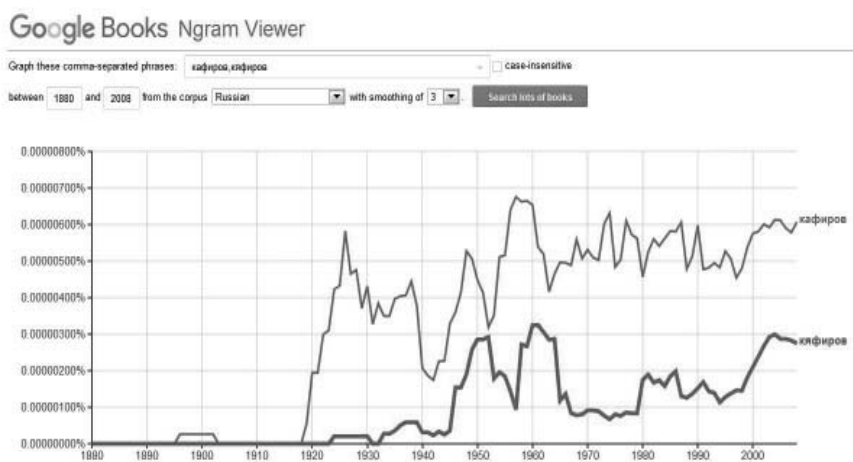


Рис. 1. Динамика употребления термина «кяфир» («кафир») в русском языке

Google Books Ngram Viewer

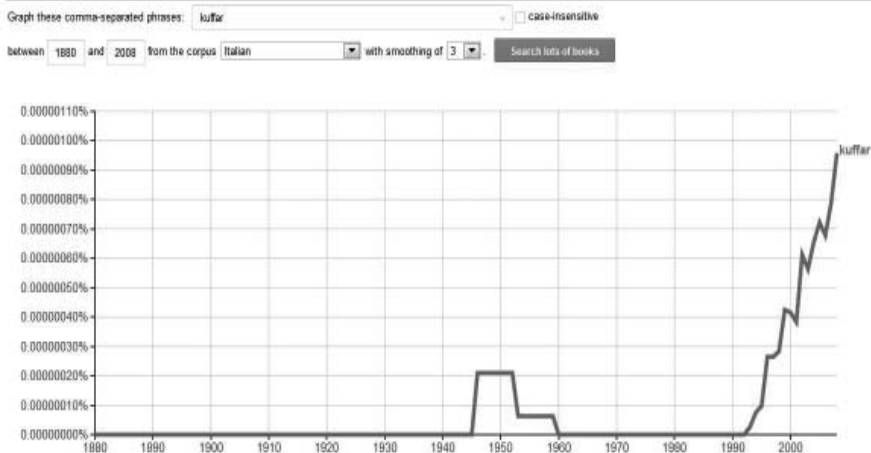


Рис. 2. Динамика употребления термина “kuffar” в итальянском языке

Google Books Ngram Viewer

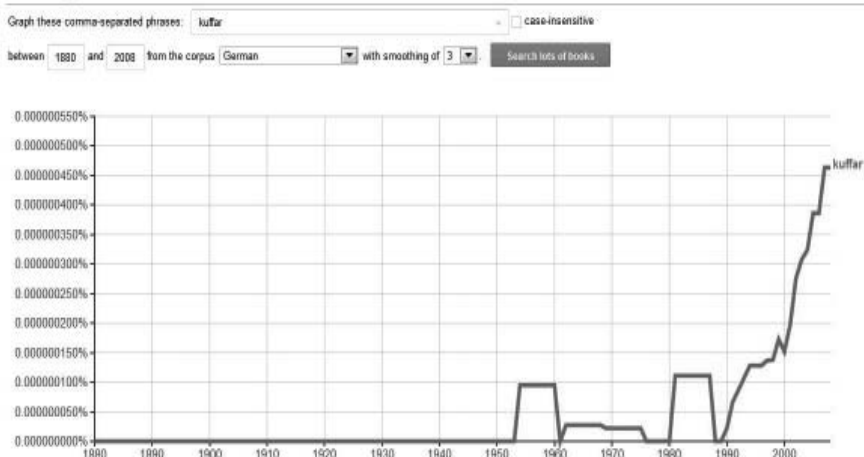


Рис. 3. Динамика употребления термина “kuffar” в немецком языке

Google Books Ngram Viewer

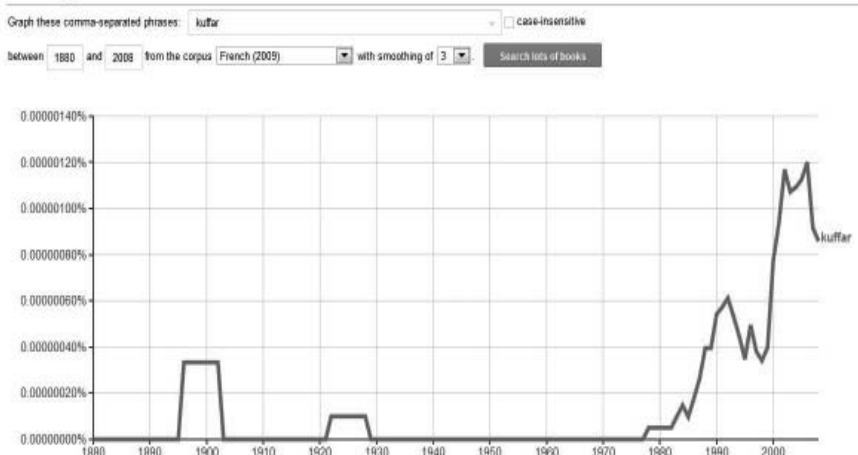


Рис. 4. Динамика употребления термина “kuffar” во французском языке

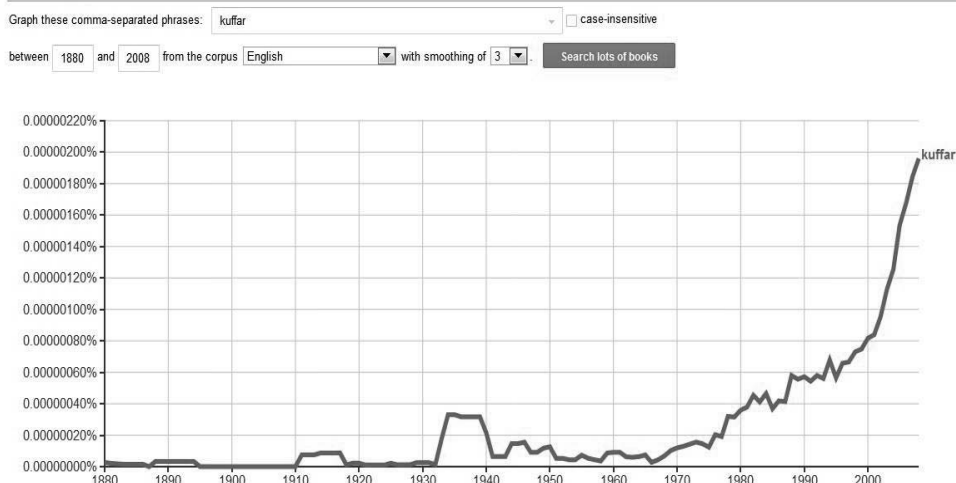


Рис. 5. Динамика употребления термина “kuffar” в английском языке

Из рисунков видно, что заметное употребление термина “kuffar” отмечалось в английском языке прежде, чем в основных европейских и русском языках. Анализ временной динамики употребления слова “kuffar” (мн.ч. от кяфир, или неверный) показал, что во французском, немецком и итальянском языках этот термин стал часто употребляться после 1980 года.

В русском языке наиболее часто термин «кяфир» употреблялся в 1960 г., а также в период 1979–1985 гг. в связи с войной в Афганистане, а затем частота употребления возросла после 2000 г. По сравнению с 1960 годом роста употребления этого слова не наблюдается.

Поскольку в английском языке стабильный рост употребления термина отмечен намного раньше (с 1965 года), чем в других языках, то можно предположить, что идеи, приводящие к росту употребления термина “kuffar” сформировались первоначально на Западе, а позже были переведены на другие языки. Динамические данные Google Ngram выявили четкий тренд увеличения частоты употребления термина «кяфир» в разных языках и позволяют делать прогноз роста его употребления в ближайшие годы. Подобные методы мониторинга и прогноза экстремистской активности на основе частоты употребления ключевых слов можно применять не только при анализе больших книжных коллекций, но также и при анализе различных источников в Интернете.

Часто ключевые слова рассматривают как неявные ссылки. Фрагменты с общими ключевыми словами можно считать – также неявно – семантически связанными.

6. Мера подобия текстов как показатель содержательного сходства

Как уже говорилось выше, неявная ссылка – это упоминание в тексте авторов, их работ и идей. Неявные ссылки представляются в виде множества значимых словосочетаний (ЗС), т.е. неявная ссылка может быть представлена одним словом/фразой/ЗС или набором близких по смыслу фраз/словосочетаний/ЗС. Разработан метод, выявляющий неявные ссылки между двумя документами, близкими по тематике. При этом неявные ссылки находятся в виде ЗС, получивших наибольшие веса, которые есть в обеих статьях, но отсутствуют в большой внешней коллекции текстов на общие темы. Весовая функция подбирается так, чтобы наиболее значимые ЗС получили наибольшие веса и попали в начало списка.

Для подхода, описанного выше, подобие текста играет важную роль. Тенденция неявного цитирования обычно простирается за пределы одного предложения.

Организация коллекции текстовых документов в виде кластеров производится таким образом, чтобы документы в пределах каждого кластера были похожи друг на друга. Текст обычно отображается в векторном пространстве, то есть документы представляются в виде «мешка слов/терминов», и каждый документ представляется как вектор признаков с использованием схемы взвешивания. Кластеризация затем выполняется путем измерения расстояния между векторами признаков.

Тематические модели основаны на идее, что документы представляют собой смеси тем, где тема является распределением вероятностей по

словам [19]. Список ключевых слов, которые представляют тему, могут быть получены с использованием этих подходов. Тем не менее, привлекаемые темы должны сопровождаться процедурой кластеризации, так как тематические модели, в принципе, не предназначены для кластеризации [14].

Авторами разработана новая мера подобия двух текстовых документов – это отношение суммы весов их общих ЗС к сумме весов ЗС в меньшем тексте. Мера подобия документа самого на себя равна 100%. В простейшем случае вес ЗС принимается равным количеству символов в ЗС. Темы могут быть описаны ключевыми словами (значимыми словосочетаниями) в данном тексте. Предложенная мера подобия была апробирована и усовершенствована на коллекции из 900 научных текстов по компьютерной графике [37]. Параметры предложенной меры подобия (такие, как минимальная учитываемая длина ЗС) рассчитывались исходя из максимума корреляции между мерой подобия пары документов и наличием между ними формальной библиографической ссылки.

Чем больше мера подобия текстов – тем выше вероятность наличия связи между этими текстами, иллюстрирующей сходство идеологических воззрений, и тем больше вероятность, что эти тексты выражают единую идеологию одной группы.

Для целей данного исследования и апробации предложенной меры подобия с помощью поисковых машин в открытом Интернете была собрана коллекция на тему «Кафиры», состоящая из 25 документов. Коллекция из 25 документов содержит $(25 \times 25 - 25) / 2 = 300$ различных пар документов. Был проведен анализ этой коллекции путем расчета меры подобия каждой пары документов. Два документа из 25 имели меру подобия 100%, поскольку один документ являлся частью другого. В данном исследовании мы считали, что пара документов связана неявной связью (неявной ссылкой), если она имеет меру подобия более 0,5%. Всего 12 пар документов имели меру подобия более 1%, и 26 пар документов имели меру подобия более 0,5%. Эти группы документов фактически являются кластерами семантически близких текстов. Так, в состав 26 пар с мерой подобия более 0,5% входят 14 различных документов. При сравнении этих документов по количеству неявных ссылок (неявно выраженных заимствований) был выявлен документ, имеющий максимальное количество (7) связей с прочими документами, который лучше всего отражает общие смысловые черты этой

группы из 25 документов, т.е. его можно считать центральным в рассматриваемом кластере документов.

Также был проведен анализ коллекции на тему «Ватники-москальи», состоящий из 31 документа. В этой коллекции были выделены два типа документов: нацистско-русофобские и их критика. Меры подобия позволили выделить помимо публицистических текстов в центре нацистско-русофобского кластера стихи и энциклопедические статьи. Использование разных жанров увеличивает спектр воздействия интертекстуальной системы документов и позволяет расширить потенциальную аудиторию. К кластеру критики относятся 7 документов. Меры подобия позволили выделить в нём центральный документ, имеющий 5 неявных связей с другими документами кластера.

7. Планируемые исследования и возможности метода

Предложенная мера подобия и кластерный анализ на основе этой меры будут полезны при анализе больших коллекций документов и для выявления скрытых связей между документами и кластерами документов.

Мера подобия будет совершенствоваться на основе принципа максимума корреляции между библиографическими и неявными ссылками в больших коллекциях научных текстов. Будут созданы новые коллекции научных текстов в различных предметных областях для более точного измерения параметров корреляции. Будут исследованы различные параметры неявных ссылок, такие, как минимальная длина значимых словосочетаний, влияющие на корреляцию.

Преобразование наблюдаемых данных об употреблении терминов в математические структуры, описывающие динамику в многолетнем временном масштабе, позволит выделить источники распространения терминов и выдвинуть предположения о формировании терминологических и идеологических траекторий.

Литература

1. Федеральный закон от 6 марта 2006 г. № 35-ФЗ «О противодействии терроризму» // СЗ РФ. – 2006. – № 11. – Ст. 1146.
2. Федеральный закон от 06.07.2016 № 375-ФЗ «О внесении изменений в Уголовный кодекс Российской Федерации и Уголовно-процессуальный кодекс Российской Федерации в части установления дополнительных мер противодействия терроризму и обеспечения общественной безопасности» // Российская газета. – 2016. – 11 июля.

3. Федеральный закон от 06.07.2016 № 374-ФЗ «О внесении изменений в Федеральный закон “О противодействии терроризму” и отдельные законодательные акты Российской Федерации в части установления дополнительных мер противодействия терроризму и обеспечения общественной безопасности» // Российская газета. – 2016. – 8 июля.
4. Указ Президента РФ от 15 февраля 2006 г. № 116 «О мерах по противодействию терроризму» (с изменениями от 02.08.2006) // СЗ РФ. – 2006. – № 8. – Ст. 897.
5. Союз «Некоммерческое партнерство высшего образования “Институт международных социально-гуманитарных связей”. Комплексный план противодействия идеологии терроризма в Российской Федерации на 2013–2018 годы» [Электронный ресурс]. – URL: http://www.imsgs.ru/students/prevention_of_extremism_and_terrorism/
6. Bakier, A.H. An Online Terrorist Training Manual – Part One: Creating a Terrorist Cell // Terrorism Focus, vol. 5, no. 13, The Jamestown Foundation, April 1, 2008. The ek-Is.org Web site has also gone under various other names, including ekhlass.org.
7. Bettencourt, L. et al. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models // Physica A. – 2006. – May. – Pp. 513–536.
8. Boucek, C. The Sakinah Campaign and Internet Counter-Radicalization in Saudi Arabia // CTC Sentinel. – 2008. – August. – P. 2. – www.carnegieendowment.org/files/CTCSentinel_Vol1Iss9.pdf.
9. Charnine, M., Kobozeva, I., Loesov, S., Schagaev, I. Graph Logic Model Framework for Predictive Linguistic Analysis. WorldComp’16, Jul. 25–28, 2016, Las-Vegas, USA [Электронный ресурс]. – URL: http://www.academia.edu/26641061/Graph_Logic_Model_Framework_for_Predictive_Linguistic_Analysis
10. Charnine, M., Klimentko, S. Measuring of “Idea-based” Influence of Scientific Papers // Proceedings of the 2015 International Conference on Information Science and Security (ICISS 2015), December 14–16, 2015. – Seoul, South Korea. – Pp. 160–164.
11. Jacobs, J.A. Receptivity Curves: Educational Research and the Flow of Ideas: Expanded Version, Population Studies Center, University of Pennsylvania, PSC Working Paper Series, PSC 13–10. – 2013. – http://repository.upenn.edu/psc_working_papers/50
12. Jee, J., Klippel, L., Hossain, M., Ramakrishnan, N., Mishra, B. Discovering the Ebb and Flow of Ideas from Text Corpora // IEEE Computer 45(2). – 2012. – Pp. 73–77.
13. How to Make Bomb in Kitchen of Mom. Featured in Al-Qaeda’s 1st English Magazine // NewsCore. July 1, 2010 [Электронный ресурс]. – URL: <http://www.foxnews.com/world/2010/07/01/make-bomb-kitchen-mom-featured-al-qaedas-st-english-magazine/>
14. Kim, S., Wilbur, W.J. Thematic clustering of text documents using an EM-based approach // Biomed Semantics. – 2012; 3(Suppl 3). – S. 6.
15. Labbe, C., Labbe, D. Detection of Hidden Intertextuality in the Scientific Publications. Dister Anne, Longree Dominique, Purnelle Gerald // 11th International Conference on Textual Data Statistical Analysis, Jun. 2012, Liege, Belgium. Liege : LAS-LA – SESLA. – 2012. – Pp. 537–551.
16. Strehl, A., Ghosh, J. Cluster Ensembles – a Knowledge Reuse Framework for Combining Multiple Partitions // Journal of Machine Learning Research. – 2002. – No. 3. – Pp. 583–617.
17. Shaparenko, B., Joachims, T. Information genealogy: Uncovering the flow of ideas in non-hyperlinked document databases // Proceedings of the Conference on Knowledge Discovery and Data Mining (KDD), 2007. – Pp. 619–628.
18. Spencer, R. Al-Qaeda newspaper: Make a bomb in the kitchen of your mom. – London : The Daily Telegraph. Archived from the original on July 4, 2010. Retrieved July 1, 2010 [Электронный ресурс]. – URL: <http://www.webcitation.org/2FAl-Qaeda-newspaper-Make-a-bomb-in-the-kitchen-of-your-mom.html&date=2010-07-04>
19. Steyvers, M., Griffiths, T. Probabilistic Topic Models. – Hillsdale, NJ, USA : Erlbaum, 2007.
20. The use of the Internet for terrorist purposes. – New York : United Nations, 2012. – 158 с. [Электронный ресурс]. – URL: http://www.unodc.org/documents/frontpage/Use_of_Internet_for_Terrorist_Purposes.pdf
21. University of Arizona. Artificial Intelligence Lab Dark Web Project. – www.icadl.org/research/terror/
22. Van Eck, N.J., Waltman, L., Dekker, R., & Van den Berg, J. A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS // Journal of the American Society for Information Science and Technology. – 2010. – No. 61 (12). – Pp. 2405–2416.
23. Weimann, G., and C. Winn. The Theater of Terror. – New York : Longman Publication, 1994.
24. Аппарат Антитеррористической комиссии Ханты-Мансийского автономного округа : методические рекомендации руководителям ор-

ганов управления образованием, образовательных учреждений и общественных организаций по вопросам противодействия распространению идеологии терроризма и вербовочной активности радикальных исламистских организаций в молодёжной среде. – Ханты-Мансийск, 2016. – 16 с. [Электронный ресурс]. – URL: www.n-var-tovsk.ru/

25. Грачев С.И., Баймашев Ю.М. Проблемные аспекты при определении понятия «терроризм» // Вестник Казанского юридического института МВД России. – 2013. – № 3 (13).

26. Грачев С.И., Товашов А.В., Завьялов А.И. Проблемы и особенности использования информационно-пропагандистского фактора в системе антитерроризма // Вестник Казанского юридического института МВД России. – 2014. – № 2 (16). – С. 62–65.

27. Ильин Е.П. Актуальные проблемы противодействия вовлечению молодежи в террористическую деятельность, 2009 [Электронный ресурс]. – URL: <http://nac.gov.ru/publikacii/stati-knigi-brosyury/ilin-ep-aktualnye-problemy-protivodeystviya.html>

28. Информационная война. Информационное противоборство: теория и практика: монография / В.М. Щекотихин, А.В. Королёв, В.В. Королёва и др.; под общ. ред. В.М. Щекотихина. – М.: Академия ФСО России, ЦАТУ, 2010. – 999 с.

29. Журавлев Д. Международный терроризм и СМИ: эволюция коммуникационного взаимодействия // Вестник РГГУ. – 2009. – № 14. Серия «Международные отношения. Регионоведение». – С. 157–169.

30. Кичерова М.Н., Кыров Д.Н. и др. Плагиат в студенческих работах: анализ сущности проблем // Интернет-журнал «Науковедение». – 2013. – № 4.

31. Котенко И.В., Юсупов Р.М. Информационные технологии для борьбы с терроризмом. Защита информации // INSIDE. – 2009. – № 2. – С. 74–79.

32. Методические материалы для проведения цикла обучающих семинаров. ФГАОУ ВО «Южный федеральный университет». Институт психологии, управления и бизнеса / под рук. П.Н. Ермакова. – Ростов н/Д., 2014. – 148 с. – С. 54–80; 85–110.

33. Методические рекомендации по совершенствованию пропагандистской работы в сфере противодействия распространению идеологии терроризма в субъектах Российской Федерации / под общ. ред. В.В. Попова. – М., 2013. – 48 с. – С. 17–28.

34. Русская правда. Халифат по законам Интернета: от twitter-революции к цифровому джихаду [Электронный ресурс]. – URL: <http://ruspravda.info/Halifat-po-zakonam-interneta-ot-twitter-revoljutsii-k-tsifrovomu-dzhihadu-13748.html>

35. Твиттер-революция [Электронный ресурс]. – URL: <http://ru.wikipedia.org/wiki/Твиттер-революция>

36. Шарнин М.М., Сомин Н.В., Кузнецов И.П., Морозова Ю.И., Галина И.В., Козеренко Е.Б. Статистические механизмы формирования ассоциативных портретов предметных областей на основе естественно-языковых текстов больших объемов для систем извлечения знаний // Информатика и её применения. ИПИ РАН – 2013. – Т. 7. – № 2. – С. 92–99.

37. Орлов Д.Е., Клименко С.В., Орлова Н.А., Шарнин М.М., Слепцов И.О. Построение и анализ тематических коллекций научных публикаций по тематике «Компьютерная графика, визуализация и виртуальная реальность» // Сборник трудов Международной конференции «Ситуационные центры и информационно-аналитические системы класса 4i для задач мониторинга и безопасности» SCVRT2015–16, 21–24 ноября 2015–2016 гг. – Пущино, Царьград, 2016. – С. 36–42.

38. Клименко С.В., Шарнин М.М., Хакимова А.Ф., Золотарев О.В., Мацкевич А.Г. Методы оценки качества и влияния (impact) научных статей для повышения объективности индекса научного цитирования // Вестник Российского нового университета. Серия «Сложные системы: модели, анализ и управление». – 2016. – Выпуск 3.

39. Золотарев О.В., Шарнин М.М., Клименко С.В. Семантический подход к анализу террористической активности в сети Интернет на основе методов тематического моделирования // Вестник Российского нового университета. Серия «Сложные системы: модели, анализ и управление». – 2016. – Выпуск 3.