

О.В. Золотарев¹
М.М. Шарнин²
С.В. Клименко³

O.V. Zolotarev
M.M. Charnin
S.V. Klimenko

**СЕМАНТИЧЕСКИЙ ПОДХОД К АНАЛИЗУ
ТЕРРОРИСТИЧЕСКОЙ АКТИВНОСТИ
В СЕТИ ИНТЕРНЕТ НА ОСНОВЕ
МЕТОДОВ ТЕМАТИЧЕСКОГО
МОДЕЛИРОВАНИЯ**

**A SEMANTIC APPROACH TO
THE ANALYSIS OF TERRORIST ACTIVITY
ON THE INTERNET BASED ON THE
METHODS OF TOPIC MODELING**

В работе исследуются вопросы анализа текстов естественного языка на предмет обнаружения неявных ссылок. В работе приведены результаты проведения семантического анализа текстов на примере анализа степени террористической угрозы в Интернете. Описываются задачи выявления террористических групп в среде Интернет. Анализируются тексты естественного языка в Интернете. Для проведения анализа текстов используется аппарат расширенных семантических сетей.

Ключевые слова: тематическое моделирование, семантическая обработка текстов, терроризм, Интернет.

The paper deals with the analysis of the degree of terrorist threats on the Internet based on the analysis of natural language texts. The article describes the tasks of identifying terrorist groups in the Internet environment. It also analyzes natural language texts on the Internet. For the analysis of texts we use the technique of extended semantic networks. The paper presents examples of the results of the semantic analysis of texts.

Keywords: topic modeling, semantic text processing, terrorism, Internet.

Введение

Данные работы выполнены в рамках грантов РФФИ:

16-07-00756 Исследование и разработка семантических методов построения «Индекса контекстного научного цитирования»;

¹ Кандидат технических наук, доцент, заведующий кафедрой информационных систем в экономике и управлении АНО ВО «Российский новый университет».

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

³ Доктор физико-математических наук, генеральный директор, АНО «Институт физико-технической информатики».

⁴ Работа выполнена при поддержке Российского фонда фундаментальных исследований, гранты 16-07-00756, 16-29-09527 и 15-07-06586. Мы благодарны РФФИ за поддержку и финансирование наших проектов.

16-29-09527 Исследования и разработка методов тематического моделирования для мониторинга, прогноза и визуализации террористической активности в информационном поле Интернета с использованием виртуального окружения;

15-07-06586 Исследование и разработка лингво-статистических методов и алгоритмов автоматического формирования ассоциативно-иерархического портрета предметной области на основе онтологий.

Интернет является не только хорошим средством повышения образовательного уровня, способствующим развитию потенциала человека, но и представляет угрозу путем распространения террористических и экстремистских взглядов. При этом необходимо различать реальный терроризм, представляющий угрозу обществу, от справедливой борьбы народа за освобождение. Интерес к проведению анализа информации в

сети Интернет на предмет выявления террористических угроз присутствует как у отдельных граждан, компаний, так и на уровне государств для обеспечения личной безопасности, защиты репутации организаций и сохранения стабильности стран. С взрывным ростом объемов информации в сети Интернет задача выявления организованных террористических группировок становится весьма актуальной в рамках предотвращения реальных угроз жизни и безопасности граждан.

Способы использования Интернета террористическими группами:

- сбор информации о предполагаемых местах терактов;
- вербовка новых членов в террористические группы;
- создание положительного имиджа террористической деятельности;
- сбор денег для поддержки террористической деятельности;
- лоббирование собственных интересов;
- вымогательство денег при угрозе кибератак;
- реклама террористической деятельности и т.д.

Рынок приложений для анализа медийного контента

Среди компаний – разработчиков подобных аналитических систем можно привести следующие.

Система SCAN предназначена для анализа медиасферы, включая социальные сети, сайты, блоги, радио, ТВ, журналы, газеты и т.д. В рамках этой системы формируются отчеты, осуществляется интеллектуальный поиск информации, реализуется визуализация связей разных типов между сообществами, компаниями, людьми, событиями... В системе строится аналитика источников информации, отслеживание динамики публикаций, поиск аффилированных компаний и персон, анализируются регионы распространения новостей, самые популярные темы. В результате анализа выделяются негативы по отраслям, строится отраслевой рейтинг, производится сравнение рейтингов компаний, выполняется анализ рисков, расчет вероятности их возникновения, формирование предупреждений о репутационных рисках и т.д.

Компания Omniture (Adobe) широко известна. Это одна из ведущих компаний на рынке аналитических. Главным продуктом компании является система сбора и обработки информации SiteCatalyst. В системе используется подход на основе меток, происходит сегментация объектов, разбиение их на группы.

Radian6 – система анализа информации в социальных сетях. В системе имеется возможность выявлять мнения людей о торговых марках, продуктах, конкурентах. Анализируются потенциальные покупатели. Компания интегрирована с Twitter и Facebook. В системе оцениваются показатели конверсии.

В Пермском государственном национальном университете разработана программа Filtext, которая отслеживает опасный контент. В системе анализируются экстремистские и оскорбительные высказывания. Анализ текстовой информации производится на нескольких уровнях – синтаксическом, морфологическом, лексическом и т.д.

В системе Techrigy/Alterian отслеживается упоминание брендов и продуктов в Интернете. Основной работы системы служит анализ ключевых слов. Система позволяет оценить положение фирмы в интернет-пространстве. Анализируются блоги, социальные сети, высказывания в сетях, анализируется аудитория с разных точек зрения, анализируется изменение отношения клиентов в исторической перспективе.

Семантический анализ интернет-контента на основе принципов построения индексов контекстного цитирования и идеологического влияния

Основным недостатком представленных выше подходов является отсутствие анализа информации в Интернете на предмет семантической близости с учетом скрытых ассоциаций. Семантический поиск информации в сети Интернет, а именно поиск объектов, их атрибутов, связей, процессов, сегодня уже не удовлетворяет растущие потребности рынка информационных аналитических приложений [27–29]. Назрела необходимость выделения не только явных отношений между объектами, но и поиска неявных ссылок, что позволит существенно расширить круг анализируемой информации с учетом скрытых ассоциаций, основанных не только на синонимии, но и на выявлении закономерностей построения преступных сообществ, их опосредованных связей, особенностей деятельности и т.д.

Индекс контекстного научного цитирования (ИКНЦ) предназначен для анализа статей и текстов в средствах массовой информации [1–3]. Основой для разработки принципов построения неявных ссылок между документами, выделенными объектами, их связями, процессами является окрестностный подход, который позволяет осуществлять поиск релевантной информации в Интернете не только по ключевым словам, но и по ассоциациям, опосредованным отношениям [4–6].

Принципы построения ИКНЦ были положены в основу разработки Индекса идеологического влияния (ИИВ) для веб-сайтов, который позволяет оценивать степень террористической угрозы от экстремистских сайтов.

Предполагается разработать методику построения динамически пополняемой базы лексических ресурсов на основе опубликованных в Сети текстовых документов (с привлечением методов корпусной лингвистики и дистрибутивной семантики). База лексических ресурсов используется в качестве источника выявления экстремистских сайтов и обнаружения семантических связей (неявных ссылок) между ними. Впервые в мировой практике на основании выявленных сетевых связей будет построен Индекс идеологического влияния сайта (ИИВ), основанный на вероятностной модели влияния (impact) идей/фраз некоторого сайта на идеи и фразы других сайтов схожей тематики. ИИВ обладает прогнозными свойствами и высокой чувствительностью, позволяющей делить новые сайты на группы и ранжировать их по степени экстремизма и влияния.

Исследование строится согласно гипотезе, что идеологическое влияние группы в Интернете способствует росту численности этой группы. При этом идеологическое влияние измеряется количеством новых идей группы, получивших распространение, а рост численности положительно коррелирует с общим количеством сайтов этой группы. В основе предлагаемого подхода лежит представление о том, что идеи могут быть адекватно выражены множеством сходных по смыслу фраз или множеством терминов (близким аналогом предлагаемого подхода выражения идей/тем является метод Скрытого распределения Дирихле – LDA). На основании этого подхода, а также на базе наблюдений за трансформациями идей во времени оказывается возможным идентификация вклада каждого сайта по отношению к выявленным темам (идеям), что, в свою очередь, позволяет обнаружить скрытые связи между сайтами/авторами.

Неявные ссылки между сайтами (упоминания похожих идей и их авторов) могут быть выявлены с помощью лингвистических и статистических методов путём поиска похожих по смыслу фраз и тем на других сайтах. Смысловое сходство фраз определяется с помощью грамматических трансформаций, программ перевода, замен синонимов на термины, полученные с помощью тематического анализа (например, методами LDA, HLDA, PLSA), а также с помощью ассоциативных связей, выявленных по автор-

ской методике построения ассоциативного портрета предметной области (АППО) [7–10].

Ведомство по патентам и товарным знакам США (USPTO) использует метод вероятностного латентного семантического анализа (PLSA) для поиска идей в патентной информации.

Трансформации получаются не только в результате замены терминов/слов на синонимы, но и на основе анализа соседних слов во фразе, которые также могут участвовать в трансформациях, т.е. могут участвовать в поисковом запросе неявных ссылок. При этом замененные слова в поисковом запросе могут не быть синонимами. Также для поиска неявных ссылок можно специальным способом подготавливать тексты, производя описанные выше замены, а после этого с использованием методов тематического моделирования выделять темы.

Методика АППО позволяет строить необходимый для анализа корпус текстов большого объёма и динамически пополнять словари террористической лексики. Вероятностная модель зависимости количества будущих неявных ссылок от количества имеющихся ссылок и их параметров строится на основе разрабатываемого в ходе исследования лингвистического процессора, выявляющего неявные ссылки. Лингвистический процессор настраивается с помощью метода машинного обучения так, чтобы корреляция между ИИВ и будущим ростом численности экстремистской группы была максимальной, а также чтобы тренды упоминания похожих идей в Интернете обладали лучшими прогнозными свойствами. Решение данной исследовательской проблемы основано на реализации комплексного подхода, сочетающего методы тематического моделирования, корпусной лингвистики, дистрибутивной семантики и визуального анализа. Подход реализуется в технологии, которая требует разработки лингво-статистических механизмов формирования ИИВ и визуального анализа его топологической структуры [21–22]. Подобная методика позволяет решать широкий круг задач как в области когнитивной семантики, так и в сфере информационного поиска. К таким задачам, например, относятся: мониторинг новых идей, оценка их влияния и эволюции во временной динамике; анализ преемственности идей [23–26].

Создание теоретических и технологических основ автоматизации построения нового показателя степени экстремизма сайта – Индекса идеологического влияния (ИИВ) – позволит прогнозировать будущие значения Индекса степени

террористической угрозы (ИСТУ), который рассчитывается на основании коллективного решения экспертов и моделируется методом машинного обучения по лексическим характеристикам сайта и прямым связям. Учитывая остроту проблемы роста экстремизма и взрывного развития Сети, актуальность предложенного исследования не вызывает сомнений. Необходимо разработать специальный лингвистический процессор, который будет находить в текстах статей неявные ссылки с заданными параметрами. Лингвистиче-

ский процессор будет разработан как модификация действующего лингвистического процессора, реализующего метод релевантных фраз, который находит похожие по смыслу фразы в других статьях и в документах из Интернета.

Необходимо также разработать базу данных для хранения тестовых коллекций текстов, представленных на экстремистских сайтах, а также информации об экспертных оценках и неявных ссылках на эти сайты. Предварительный вариант схемы данных представлен на рис. 1.

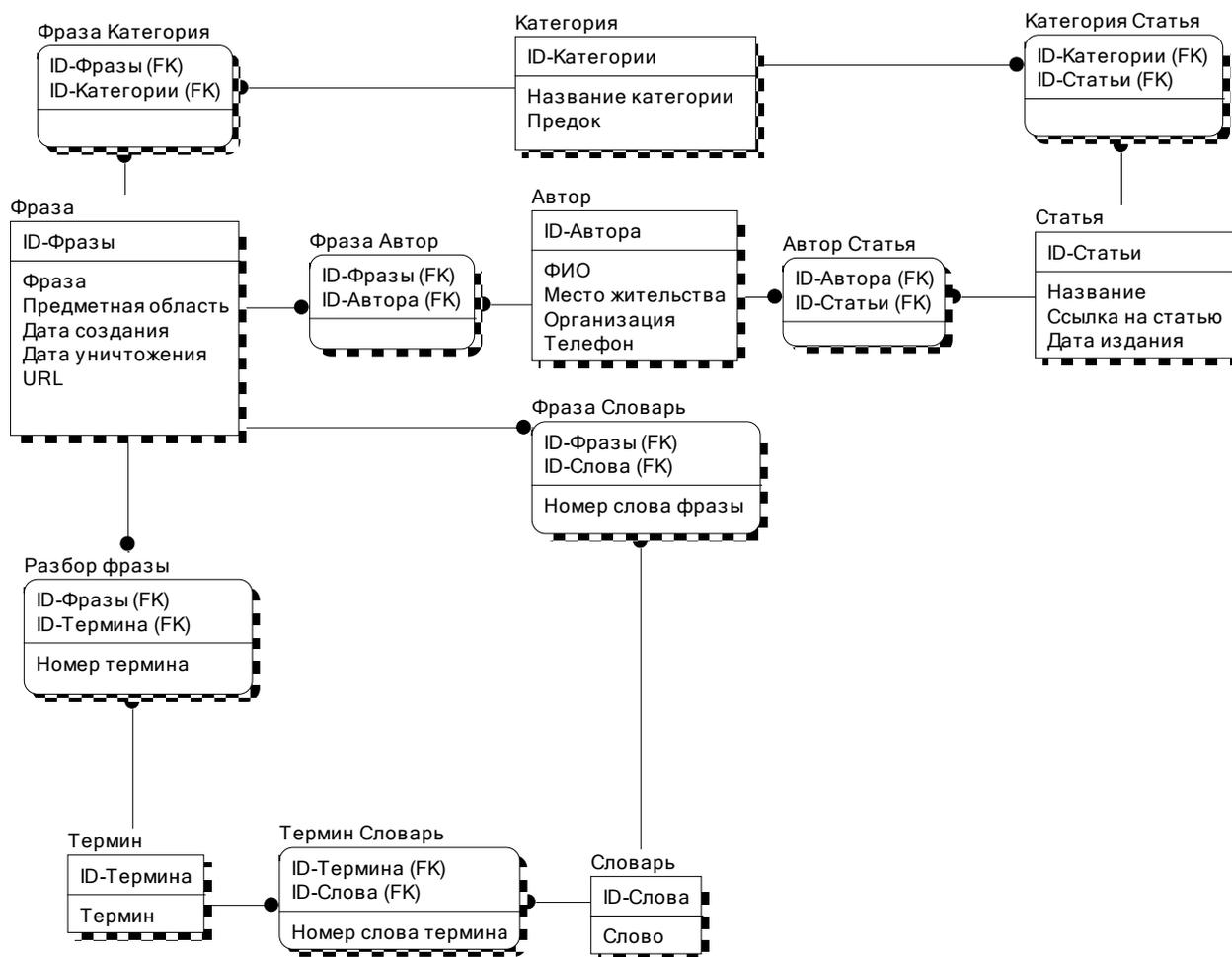


Рис. 1. Схема данных

Предполагается выделять темы, характерные для конкретной предметной области [29–31]. Тема характеризуется набором признаков. При этом один документ может относиться к нескольким предметным областям, соответственно, содержать несколько тем, каждая из которых, в свою очередь, может относиться к нескольким предметным областям.

Будут разработаны программы расчета индексов ИСТУ и ИИБ и программы расчета корреляции этих индексов.

Следует иметь в виду, что неявных ссылок гораздо больше, чем экспертных оценок, т.к. во-первых, не всегда упоминание чужих экстремистских идей и их авторов в текстах сайтов сопровождается изменением соответствующих

экспертных оценок, и, во-вторых, похожие идеи, но полученные самостоятельно, также редко сопровождаются изменением экспертных оценок соответствующих сайтов с подобными идеями. В силу большего количества неявных ссылок индекс ИИВ имеет большую чувствительность и количество различных значений, чем ИСТУ. Особенно это важно для реальных баз данных по терроризму, имеющих огромное количество новых документов/текстов, для которых еще не сделаны экспертные оценки, среди которых необходимо автоматически выделить тексты с наибольшей степенью террористической угрозы и отсеять менее правдоподобные угрозы с более низкой вероятностью.

Исследование параметров неявных ссылок будет базироваться на гипотезе о том, что сайты с новыми экстремистскими идеями, на которые много неявных ссылок, имеют повышенную вероятность террористической угрозы и более высокие значения индекса ИСТУ. В частности будут исследованы оптимальные частотные и временные характеристики неявных ссылок, при которых достигается максимум корреляции индексов ИИВ и ИСТУ.

Вероятностная математическая модель зависимости количества экспертных подтверждений ИСТУ от количества неявных ссылок и их параметров будет строиться на основе лингвистического процессора, выявляющего неявные ссылки, который настраивается с помощью метода машинного обучения так, чтобы корреляция между индексами ИСТУ и ИИВ была максимальной.

Сбор и обработка данных с использованием средств семантического анализа

Предлагаемые методы и подходы являются оригинальными, экономически эффективными и патентоспособными. Указанные методы и подходы основаны на оригинальных разработках коллектива исполнителей.

Предлагаемые в проекте методы и подходы являются инновационными, т.к. они позволяют не только обнаруживать и выделять целевую информацию из интернет-пространства, но и обнаруживать распределения и зависимости предметных (идеологических) траекторий [32].

Существующие методы оценки значимости текстов могут высоко оценить только те документы (сайты), на которые имеются прямые ссылки или гиперссылки. В настоящее время такие документы (сайты) составляют лишь очень малую долю от общего объема документов, причем последний растет быстрее, чем количество соответствующих ссылок. С этой проблемой

особенно остро сталкиваются владельцы новых веб-сайтов, страдающие от недостатка ссылок и интернет-трафика.

Предлагаемые методы и подходы решают эту проблему, т.к. предлагают оценивать степень экстремизма и идеологического влияния сайтов по наличию в них информативных фраз, которые образуют неявные контекстные ссылки между сайтами [8–13].

Семантический анализ больших коллекций неструктурированного текста является актуальной проблемой. Тематическое моделирование – одно из приложений машинного обучения к анализу текстов, которое определяет тематическую структуру каждого документа из текстовой коллекции, а также тематический профиль всех слов этой коллекции. Современный анализ текстов практически всегда работает с большими объемами данных, которые можно обработать лишь с помощью подобных алгоритмов тематического моделирования [31, 32].

Информативные фразы можно найти с помощью метода релевантных информативных фраз [4–7], суть которого состоит в поиске наиболее информативных фрагментов (предложений, частей предложений, отдельных терминов) в текстах, принадлежащих определенной предметной области. Предметная область задается множеством ключевых терминов с дальнейшим переходом к релевантным этому множеству текстовым документам и формированию ее ассоциативного портрета (М.М. Шарнин) [1–3; 11].

Оценка информативности фразы выполняется путем подсчета индекса встречаемости этой фразы и ее семантических эквивалентов в других документах. Рейтинг/значимость сайта вычисляется с учетом входящих в него информативных фраз, количества прямых ссылок (гиперссылок), неявных контекстных ссылок [33; 34].

В качестве метода сбора данных из социальных сетей в настоящем проекте предлагается использовать систему IQBuzz. Руководитель данного исследования – М.М. Шарнин [12–16].

В качестве метода сбора данных из Интернета предлагается использовать KeyCrawler [15–17; 4; 5] – оригинальный семантический навигатор по Интернету, разработанный и развиваемый данным коллективом авторов, который обеспечивает извлечение информационных объектов по исходным семантическим образам – наборам ключевых и опорных терминов. KeyCrawler был успешно частично апробирован в энциклопедии ключевых понятий KEYWEN [7–9; 4; 17], осуществляющей направленное извлечение энциклопедической информации из Интернета. Дан-

ные из открытого Интернета являются ценным дополнением к информации из IQBuzz, т.к. по предварительным оценкам количество документов в открытом Интернете для некоторых запросов на русском языке в настоящее время больше в несколько тысяч раз, чем количество аналогичных документов в IQBuzz.

В исследовании также предлагается использовать метод машинного обучения и корреляционный анализ для определения оптимальных параметров неявных ссылок, а также методы коллективного интеллекта и краудсорсинга [18–21] для оценки качества алгоритмов выявления неявных ссылок и определения смысловой схожести фраз, представляющих общую идею.

Заключение

Предложенные решения представляют комплексный подход к автоматической обработке текстовых документов на основе сочетания методов тематического анализа, машинного обучения, статистики, корпусной лингвистики и дистрибутивной семантики с целью выявления неявных контекстных ссылок, информативных фраз и их ассоциативных связей [18–20]. Проект реализуется на сверхбольших объемах данных (Big Data), представленных в среде Интернета и в социальных сетях.

Предложенные методы обработки текстов в Интернете позволяют выявлять террористически настроенные сайты на основе отслеживания динамики изменения Индекса степени террористической угрозы (ИСТУ), который в дальнейшем будет в существенной степени формироваться в автоматическом режиме, что позволит решать вопрос о степени террористической угрозы, исходящей от интернет-сообществ, сайтов и отдельных индивидуумов в режиме онлайн. Кроме этого, степень изменения индекса ИСТУ позволит строить прогнозы об актуальности отдельных идей, о расширении или сокращении деятельности террористических группировок.

Литература

1. Charnine, M., Klimenko, 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. – Seoul, South Korea, 2015. – Pp. 160–164.
2. Charnine, M., Somin, N., Klimenko, S., Ezhe-la, V. Linguistic Approach to Scientometrics // Proceedings of the 2015 International Conference on Artificial Intelligence (ICAI 2015), WORLD-COMP’15, July 27–30, 2015. – Las Vegas, Nevada, USA. – V. II. – Pp. 812–817.

3. Ежела В.В., Клименко С.В., Райков А.Н., Шарнин М.М. Поиск альтернативных подходов к построению оценки качества научных публикаций // Труды Международной конференции и Школы по физико-технической информатике СРТ2014, 11–18 мая 2014 г., Ларнака, Республика Кипр; 17–21 июня 2014 г., Протвино, Московская область. – Протвино–Москва: Изд. ИФТИ, 2015. – С. 53–61.

4. Charnine, M., Somin N., Nikolaev, V. Conceptual Text Generation Based on Key Phrases // Proceedings of the 2014 International Conference on Artificial Intelligence (ICAI 2014). – Vol. I, WORLDCOMP’14, July 21–24, 2014. – Las Vegas Nevada, USA: CSREA Press, 2014. – Pp. 639–643.

5. Михеев М.Ю., Сомин Н.В., Галина И.В., Золотарев О.В., Козеренко Е.Б., Морозова Ю.И., Шарнин М.М. Фальштексты: классификация и методы опознания текстовых имитаций и документов с подменой авторства // Информатика и ее применения. – 2014. – Т. 8. – Вып. 4. – С. 70–77.

6. Шарнин М.М., Сомин Н.В., Галина И.В., Родина И.В., Николаев В.Г. Осмысленные генеранты: метод порождения текстов повышенной информативности из релевантных фраз // Труды Международной конференции и Школы по физико-технической информатике СРТ 2014, 11–18 мая 2014 г., Ларнака, Республика Кипр; 17–21 июня 2014 г., Протвино, Московская область. – Протвино–Москва: Изд. ИФТИ, 2015. – С. 87–91.

7. Charnine, M. Keywen: Automated Writing Tools. – USA: Booktango, 2013.

8. Charnine, M., Charnine, V. Keywen Category Structure : монография. – USA: Wordclay, 2008. – Pp. 1–60.

9. Шарнин М.М., Кузнецов И.П. Автоматическое формирование электронных энциклопедий и справочных пособий по информации из сети Интернет // Системы и средства информатики. ИПИ РАН. – 2004. – Вып. 14. – С. 210–223.

10. Galina, I., Charnine, M., Somin, N., Nikolaev, V., Morozova, Yu., Zolotarev, O. Metod for generating subject area associative portraits: different examples // Proceedings of the 2015 International Conference on Artificial Intelligence (ICAI 2015), WORLDCOMP’15, July 27–30, 2015. – Las Vegas, Nevada, USA, 2015. – V. I. – Pp. 288–294.

11. 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. – 2013. – Т. 7. – № 2. – ИПИ РАН. – С. 92–99.

12. Charnine, M., Petrov, A., Kuznetsov, I. Association-Based Identification of Internet User Interests // Proceedings of the 2013 International Conference on Artificial Intelligence (ICAI 2013). – V. I., WORLDCOMP'13. July 22–25, 2013. – Las Vegas, Nevada, USA: CSREA Press, 2013. – Pp. 77–81.

13. Zolotarev, O., Charnine, M., Matskevich, A. A Conceptual Business Process Structuring by Extracting Knowledge from Natural Language Texts // Proceedings of the 2014 International Conference on Artificial Intelligence (ICAI 2014). – Vol. I, WORLDCOMP'14, July 21–24, 2014. – Las Vegas, Nevada, USA: CSREA Press, 2014. – Pp. 82–87.

14. Kuznetsov, I.P., Kozerenko, E.B., Charnine, M.M. Technological peculiarity of knowledge extraction for logical-analytical systems // Proceedings of the 2012 International Conference on Artificial Intelligence, ICAI 2012. – Vol. 2. – Las Vegas, USA, 2012. – Pp. 762–768.

15. Kuznetsov, I., Charnine, M., Kozerenko, E., Somin, N., Nikolaev, V., Matskevich, A. Intelligent tools for the semantic Internet navigator design // CEUR Workshop Proceedings. – 2012. – Vol. 934. – Pp. 222–231.

16. Морозова Ю.И., Козеренко Е.Б., Будзко В.И., Кузнецов К.И., Шарнин М.М. Семантическая структуризация текстовых знаний для систем аналитического мониторинга больших объемов информации в социальной сфере // Системы высокой доступности. – 2014. – № 3. – С. 21–34.

17. Charnine, M.M., Kuznetsov, I.P., Kozerenko, E.B. Semantic Navigator for Internet Search Proceeding of International Conference on Machine Learning, 27–30, 2005. – Las Vegas, USA: CSREA Press, 2005. – Pp. 60–65.

18. Protasov, V., Charnine, M., Melnikov, E. The Crowdsourcing Linguistic Technology for Experts Assessment // Proceedings of the 2014 International Conference on Artificial Intelligence (ICAI 2014). – Vol. I, WORLDCOMP'14, July 21–24, 2014. – Las Vegas, Nevada, USA: CSREA Press, 2014. – Pp. 656–661.

19. Charnine, M., Protasov, V. Optimal Automated Method for Collaborative Development of University Curricula // Proceedings of the 2013 International Conference on Artificial Intelligence (ICAI 2013). – V. I. WORLDCOMP'13. July 22–25, 2013. – Las Vegas, Nevada, USA: CSREA Press, 2013. – Pp. 96–100.

20. Шарнин М.М., Шагаев И., Протасов В.И., Родина И.В., Золотарев О.В., Попова О.А. Ис-

пользование веб-семантики для совершенствования образовательных программ вузов // Вестник МГГУ им. М.А. Шолохова. Филологические науки. – 2015. – № 2. – С. 97–112.

21. Bacon, Elisabeth, Hagel, George, Charnine, Michael, Foggie, Richard, Kirk, Brian, Schagaev, Igor, Kravtsov, George. WEDUCA: Web-enhanced design of university curricula // Proceedings of the FECS-13: The International Conference on Frontiers in Education: Computer Science and Computer Engineering, July 22–25, 2013. – Las Vegas, Nevada, USA: CSREA Press, 2013. – Pp. 288–294.

22. Золотарев О.В., Козеренко Е.Б., Шарнин М.М. Проведение аналитической разведки на основе анализа неструктурированной информации из различных источников, включая Интернет и средства массовой информации // Вестник Российского нового университета. Серия «Сложные системы: модели, анализ и управление». – 2015. – Выпуск 1. – С. 49–54.

23. Шарнин М.М., Золотарев О.В., Сомин Н.В. Извлечение и обработка знаний из неструктурированных текстов деловой сферы и социальных сетей // 4-я Международная научно-практическая конференция «Социальный компьютеринг: основы, технологии развития, социально-гуманитарные эффекты». – М. : МПГУ, 2015.

24. Zolotarev, O.V., Charnine, M.M., Matskevich, A.G., Kuznetsov, K.I. Business Intelligence Processing on the Base of Unstructured Information Analysis from Different Sources Including Mass Media and Internet // Proceedings of the 2015 International Conference on Artificial Intelligence (ICAI 2015). – Vol. I, WORLDCOMP'15, July 27–30, 2015. – Las Vegas, Nevada, USA, 2015. – V. I. – Pp. 295–299.

25. Золотарев О.В., Шарнин М.М. Методы извлечения знаний из текстов естественного языка и построение моделей бизнес-процессов на основе выделения процессов, объектов, их связей и характеристик труды Международной научной конференции СРТ 2014 Международная научная конференция Московского физико-технического института (государственного университета), Института физико-технической информатики. – М. : Институт физико-технической информатики, 2015. – С. 92–98.

26. Золотарев О.В. Козеренко Е.Б., Шарнин М.М. Принципы построения моделей бизнес-процессов предметной области на основе обработки текстов естественного языка // Вестник Российского нового университета. – 2014. – Выпуск 4. – С. 82–88.

27. Золотарев О.В. Процессный подход к управлению в проектах внедрения корпоративных информационных систем. // Вестник Российского нового университета. – 2014. – Выпуск 4. – С. 89–92.

28. Золотарев О.В. Методы выделения процессов, объектов, отношений из текстов естественного языка // Проблемы безопасности российского общества. – Смоленск : Свиток. – 2014. – № 3–4. – С. 276–283.

29. Золотарев О.В. Инновационные решения в формировании функциональной структуры предметной области // Вестник Российского нового университета. – 2013. – Выпуск 4. – С. 82–84.

30. Золотарев О.В. Методы и инструменты моделирования предметной области // Труды Тринадцатой Международной научной конференции «Цивилизация знаний: проблемы и перспективы социальных коммуникаций» – М. : РосНОУ, 2012.

31. Золотарев О.В. Новые подходы в постро-

ении функциональной структуры предметной области // Труды по материалам конференции «20 лет постсоветской России: кризисные явления и механизмы модернизации». – Екатеринбург : Гуманитарный университет, 2011.

32. Золотарев О.В. Формализация знаний о предметной области на основе анализа естественно-языковых структур // Труды Двенадцатой Международной научной конференции «Цивилизация знаний: проблема человека в науке XXI века». – М. : РосНОУ, 2011.

33. Золотарев О.В. Использование моделей представления знаний в образовании // Труды по материалам конференции «Образование в XXI веке. Проблемы и перспективы». – Пенза : ПДЗ, 2010.

34. Золотарев О.В. Средства анализа информации в системах, основанных на семантических сетях // Труды Одиннадцатой Международной научной конференции «Цивилизация знаний: проблемы модернизации России». – М. : РосНОУ, 2010.