

Актаева А.¹, Галиева Н.², Асанова К.³, Наралиев Н.⁴ Сундетов С.⁵, Макулбек Н.⁶

¹Казахская академия транспорта и коммуникации им. М.Тынышбаева, г.Алматы, Казахстан

^{2,4}Павлодарский государственный университет им. С. Торайгырова, г.Павлодар, Казахстан

³Казахско-американский университет, г.Алматы, Казахстан

^{5,6}Казахская академия транспорта и коммуникаций им. М.Тынышбаева, г.Алматы, Казахстан

МЕТОДИКА ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ НА ОСНОВЕ ОНТОЛОГИИ

АННОТАЦИЯ

В статье обсуждается значимость в современном информационном обществе информационных ресурсов. Рассматривается структура и основные принципы технологии повышения вероятности идентификации субъектов информационных процессов открытых ресурсов сети Интернет на основе методов онтологии.

КЛЮЧЕВЫЕ СЛОВА

Социальная сеть; Интернет; Web-документ; База данных; язык SPARQL; онтология; семантика.

Aktaeva A.¹, Galieva N.², Asanova K.³, N. Naraliev⁴, S. Sundetov⁵, N. Makulbek⁶

¹Kazakh Academy of Transport and Communication named after M. Tynyshpaev, Almaty, Kazakhstan

^{2,4}S. Toraihyrov Pavlodar State University, Pavlodar, Kazakhstan

³Kazakh-American University, Almaty, Kazakhstan

^{5,6}Kazakh Academy of Transport and Communication named after M. Tynyshpaev, Almaty, Kazakhstan

TECHNIQUE OF IDENTIFICATION OF USERS OF SOCIAL NETWORKS OVER ONTOLOGIES

ABSTRACT

In the article the importance of information resources in today's information society is discussed. The structure and the basic principles of the technology on ontologies increase the probability of identifying the subjects of information processes of open Internet resources of its work is considered.

KEYWORDS

Social Network; Internet; Web-document; database; the SPARQL language; ontology; semantics.

Одним из приоритетов модернизации экономики страны является развитие стратегических информационных технологий, являющихся наиболее важной составляющей процесса информатизации общества.

Технический прогресс меняет ИКТ ускоренными темпами. К настоящему времени этап их развития определяется появлением все новых технологических средств поиска и переработки данных. Сеть Интернет занимает среди них одно из главенствующих мест. Интернет развивается столь стремительно, что ежегодно происходят его качественные изменения.

А появившиеся не так давно «облачные» вычисления оказывают все большее влияние на многие направления развития технологии и бизнеса.

Новейшие Интернет-технологии теперь широко применяются в различных областях науки, техники и бизнеса и стремительно перешли в все сферы человеческого бытия: бытовая сфера – умный дом, бытовые приборы с помощью средств Интернет-технологий; средства связи – IP- телефония, IP-телевидение; гаджеты – сотовые телефоны, и др. [2,15].

Знание динамики и особенностей процессов развития процессов информатизации общества необходимое условие для принятия мер по наиболее полной информатизации общества, увеличения числа Интернет-пользователей, совершенствования качества онлайн-услуг, в том числе и предоставляемых государством, продвижения отечественных программных продуктов на мировом рынке. Это, в свою очередь, обуславливает необходимость контроля субъектов информационных процессов для идентификации возможных направлений информационного

воздействия на пользователей сети интернет. В связи с тем, что порядка 80% информации в сети Интернет представлено в текстовом виде, возникает необходимость в разработке средств идентификации пользователей на основе методов математической лингвистики и семантики, предназначенных для обеспечения информационной безопасности важных объектов политической, социально-экономической, оборонной, культурной и других сфер деятельности от внешних и внутренних угроз модификации информации [2,15,14].

Технологические элементы системы модификации информации- манипуляторное воздействие на пользователей социальных сетей. Для этого используется комплексный критерий, который основывается на учете сочетания следующих параметров:

1. Частота использования технологий;
2. Широта их применения;
3. Степень воздействия.

С учетом данного комплексного критерия можно выделить группы манипуляторных приемов, имеющих наиболее универсальный характер, с высокой частотой встречаемости в различных технологиях информационных войн, имеющих сферу применения в таких информационно-коммуникативных ситуациях, как публичные дискуссии и групповые обсуждения, выступления на митингах и демонстрациях, в средствах массовой коммуникации, в межгрупповых и межличностных конфликтах, характеризующиеся достаточно высокой действенностью и влиянием на психику человека. Данные приемы характеризуются высокой степенью выраженности по всем трем параметрам манипуляторного воздействия на пользователей социальных сетей.

Применение методов обработки текстовых сообщений для защиты от модификации информации крайне важно в случаях, когда однозначная идентификация при помощи вычисления сетевых данных, таких методов идентификации, использующие технические характеристики, такие как:

1. HTTP Cookie;
2. IP-адрес;
3. MAC-адрес;
4. геолокационные данные;
5. данные об используемой ОС, браузере, параметрах оборудования (разрешение и размер экрана, центральный процессор) и т.д., о компьютере пользователя невозможна, т.к. компьютером могут пользоваться несколько человек, или компьютер может находиться в некоторой локальной сети.

Большое количество интернет ресурсов и сервисов, таких как форумы, порталы, интернет-магазины, сталкиваются с различными проявлениями проблемы манипуляции и искусственного формирования общественного мнения, путем «организации» целенаправленных тематических диалогов, в которых ряд пользователей имеют несколько учетных записей.

Приемы информационно-психологического воздействия на массовое сознание выделено на следующие семь основных групп информационно-психологического воздействия:

1. «Приклеивание или навешивание ярлыков» (name calling);
2. «Сияющие обобщения» или «блистательная неопределенность» (glittering generality);
3. «Перенос» или «трансфер» (transfer);
4. «Ссылка на авторитеты», «по рекомендации», «свидетельства» или «свидетельствование» (testimonial);
5. «Свои ребята» или «игра в простонародность» (plain folks); «Перетасовка» или «подтасовка карт» (card stacking);
6. «Общая платформа» или «фургон с оркестром» (band wagon) [2,15].

Возможность использования порталов и сайтов для распространения информации и недостаточная функциональность механизмов идентификации и аутентификации пользователей, оставляющих сообщения, определяет ряд направлений совершенствования систем защиты и систем мониторинга информационной безопасности ИТКС.

Задача выделения информации (information extraction) отличается от задачи поиска информации (information retrieval). Системы выделения информации можно разделить на следующие четыре типа по степени участия эксперта в создании и адаптации системы:

1. *Настраиваемые вручную*, где Пользователь задает на некотором языке правила выделения информации из конкретных сайтов.
2. *С обучением*. Пользователь вручную размечает обучающую выборку документов, которая используется для построения модуля выделения информации.

3. *С частичным обучением.* Пользователь не размечает всю обучающую выборку, а лишь предоставляет некоторую дополнительную информацию, например, выбирает шаблон из вариантов, представленных системой, и отмечает данные, которые нужно выделить.

4. *Без обучения.* Система автоматически размечает обучающую выборку и создает модуль выделения информации полностью без участия пользователя [10,11,13].

Поиск информации заключается в выборе релевантных документов из крупной коллекции. В рамках этой задачи документ представляет собой ни что иное, как набор слов, значения и связи которых не рассматриваются. Поисковая система в общем случае не позволяет выполнять сложные аналитические запросы, требующие анализа содержания документов. Направление извлечения данных, которое является менее разработанным, как раз ставит задачу выделения структуры, то есть значения информации из неструктурированных текстов.

Задача выделения информации из неструктурированного текста является важной на направлении обработки и анализа данных. Актуальность задачи обусловлена быстро возрастающим объемом неструктурированной информации, например, в сети Интернет. В общем случае выделение информации подразумевает заполнение некоторой структуры данных, или шаблона, информацией, содержащейся в текстовых данных на естественном языке, или, другими словами, выявление экземпляров определенных классов объектов или событий и связей между ними. Согласно исследованию Рассела и Норвига, выделение информации находится «посередине» между поиском информации (information retrieval), который заключается в подборе документов, отвечающих запросу пользователя, и пониманием смысла текста, подразумевающим глубокий анализ текста с целью выявления его семантики [1, 3, 4, 10,11,13].

В последнее время все большее внимание исследователей обращается к задаче выделения информации из Всемирной паутины. HTML-страницы в Интернет обладают тем преимуществом перед обычными текстами, что в них присутствуют элементы разметки, например, списки, заголовки и таблицы.

Другой особенностью Web-документов является тот факт, что информация в них обычно генерируется автоматически из некоторых баз данных, что превращает процесс выделения информации из таких текстов в процесс «декодирования». Наконец, нельзя не отметить огромное количество данных, содержащихся в Интернет и их неоднородность.

Перечисленные особенности обуславливают набор методов, обычно применяемых при выделении данных из Всемирной паутины. Если в традиционных системах выделения информации из неструктурированных текстов обычно применяются методы обработки естественного языка, такие как словари и грамматики, то в системах выделения информации из Интернет чаще используются алгоритмы машинного обучения и выделения шаблонов, которые опираются на синтаксические свойства и визуальную структуру Web-страниц (рис.1).



Рис. 1. Схема функционирования информационного объекта при работе с пользователями

Информация на Web-страницах в Интернет делится на два класса:

1. Информация, которая генерируется автоматически из структурированных БД;
2. Информация, публикуемая вручную.

Для выделения информации может использоваться обучение на подготовленной выборке источников, и/или эвристические модели, которые могут быть основаны, например, на использовании заданных лексико-синтаксических шаблонов или же на онтологиях. Направление извлечения информации с помощью онтологий выделилось из общей задачи выделения

информации сравнительно недавно, но уже отмечено как перспективное направление развития систем выделения информации.

Представление знаний с помощью онтологий в информационно-поисковых системах позволяет выполнять сложные структурированные (объектно-ориентированные) запросы, которые можно представить в виде графа. Практически выполнение такого запроса сводится к поиску подграфа в графе. На основе онтологии информационно-поисковая система может предложить конкретизировать или, наоборот, расширить запрос, если ему удовлетворяет слишком много (соответственно, слишком мало) объектов. Более того, онтология позволяет предложить пользователю самому осуществить обзор области знания с помощью навигации по понятиям, переходя от одного понятия к другому по связям между ними.

Во многих системах, основанных на онтологиях, используются следующие методы извлечения информации:

- Языковые правила, сформулированные с помощью регулярных выражений;
- Справочники (gazetteers);
- Методы классификации;
- Построение частичных деревьев синтаксического разбора предложений с семантической аннотацией;
- Анализ HTML/XML-тегов;
- Поиск в Интернет;
- Описание используемых алгоритмов.

Отметим еще несколько преимуществ использования онтологий для представления знаний:

- Гибкость модели данных, которая позволяет сравнительно легко ее изменять и расширять.
- Возможность повторного применения существующих онтологий.

Использование онтологий является естественным развитием направления. Онтология – это «формальная, явная спецификация общей концептуализации». Другими словами, это способ представления знаний с помощью конечного множества понятий и отношений между ними. Онтологии используются для формального описания области знания. Понятие, или сущность – это класс индивидуальных объектов, или экземпляров. Связи между понятиями бывают следующих типов:

- иерархические;
- свойства;
- ограничения значений
- определяющие непересекаемость понятий;
- конкретизирующие логические отношения.

Для задания онтологии предметной области используются следующие классы языков, основанных на различных математических моделях:

1. Языки, основанные на графических моделях.
2. Языки, основанные на математической логике.
3. Языки, основанные на вероятностных и нечетких моделях.

Использование онтологий для представления знаний имеет ряд преимуществ:

1. Онтология содержит общее видение области знания, что позволяет устранить различия в терминологии, проблемы синонимии и многоязычности [10,11,13].

2. Математический аппарат дескриптивной логики предоставляет средства логического вывода новых фактов на основе существующих. Например, *математическая модель алгоритма Brainstern*: пусть W – множество всех слов, которые встречаются во всех документах заданной коллекции Doc, включая \square – пустое слово, а PW – множество всех упорядоченных пар слов, то есть $PW = W \times W$. Определим документ d как отображение $d: N \rightarrow W$, которое сопоставляет каждому натуральному числу n слово, стоящее на n -той позиции в данном документе коллекции. Номера позиций, на которых нет слов (после конца документа), отображаются в пустое слово. Аналогично определим абзац p как отображение $p: N \rightarrow W$, которое сопоставляет каждому натуральному числу n слово, стоящее на n -той позиции в данном абзаце. Номера позиций, на которых нет слов, отображаются в пустое слово. Обозначим множество всех абзацев в коллекции через P . Определим рубрику r как произвольное подмножество множества документов, а именно – $r \in 2^{doc}$. Мощность рубрики, как количество документов в ней, будем обозначать через $|r|$. Обозначим множество всех заданных рубрик через R [10,11,13].

Таким образом, использование онтологий обеспечивает выполнение требования к разработанной системе, которое заключается в необходимости осуществления логического вывода новых знаний.

Выделение информации традиционно имеет целью найти сведения, которые описывают некоторую область знаний, заданную структурой данных. Онтологии же как раз и представляют собой формальную модель предметной области, выраженную, например, в виде графа понятий и связей, что обобщает иерархическую структуру данных, обычно используемую для заполнения в задаче выделения информации. Во многих системах выделение информации включает в себя следующие этапы:

- 1) поиск информации;
- 2) выделение терминов;
- 3) выделение именных групп, например, имен людей, организаций, географических мест;
- 4) выделение слов и словосочетаний, которые обозначают один и тот же объект (coreference resolution);
- 5) нормализация терминов, которая позволяет связать их с формализованным описанием предметной области;
- 6) выделение семантических отношений между терминами;
- 7) поиск дублирующих записей;
- 8) нормализация записей, то есть приведение их к стандартному виду.

Логический вывод новых утверждений позволяет превращать скрытое знание в явное, а также находить противоречия в онтологиях. Следует отметить, что неявное знание, как правило, и содержит смысл явного. Именно в этом контексте принято утверждать, что в рамках семантического подхода компьютер «понимает» смысл информации. Задание специальных связей (таких, как «является подклассом») между понятиями в терминах логического вывода по существу означает задание семантики области знания. С помощью логического вывода можно автоматически классифицировать экземпляры по классам на основе их атрибутов. Большим достоинством логического вывода является возможность не вводить в онтологию всю информацию, а ввести лишь ее базовую часть, из которой можно вывести все необходимые утверждения. Заметим, что правила логического вывода отличаются для различных языков на основе RDF-модели, так как эти языки имеют разную семантику.

При использовании РБД для объединения информации необходимо согласование множества величин, которые имеют технический характер, например, названий полей таблиц. Онтологии позволяют вынести процесс согласования на семантический уровень. Это означает, что если разные организации используют одну и ту же онтологию, то есть модель данных, то конкретная реализация системы ее хранения не имеет значения. Объединение данных в таком случае можно осуществить со сравнительно небольшими затратами ресурсов, просто соединив две базы утверждений-троек, входящих в фактологические части онтологий.



Рис.2. Общая архитектура АИСУ информацией [7]

Лингвистический подход подразумевает выделение ключевых слов и фраз на основе морфологии и семантики текста. В рамках этого подхода широко используются базы семантических связей между словами, словари синонимов, средства морфологического разбора и анализа. Самые известные разработки на основе лингвистического подхода – это системы TERMS и LEXTER. Основным проблемным вопросом при реализации лингвистических методов является «шум», а именно – большое количество словосочетаний, характерных для естественного языка, но не являющихся терминами [7].

Статистический подход подразумевает оценивание важности слова на основе информации о числе его вхождений в документы. Для этого чаще всего используются модели на основе TF-IDF – нормализованного веса слова. В рамках статистического подхода широко применяются средства, основанные на вероятностных моделях [7].

С одной стороны, это позволяет использовать мощный математический аппарат и получать теоретические доказательства эффективности алгоритмов. С другой стороны, большинство таких подходов имеют один существенный недостаток. Его суть в том, что они опираются на предположение о независимости появления слов в предложении. Для естественного языка такая гипотеза является слишком сильным условием [7].

В отличие от лингвистического подхода, основным проблемным вопросом при реализации статистического подхода является «*тишина*» – ситуация, при которой термины, состоящие из одного слова, остаются незамеченными системой. В исследованиях последних лет статистические и лингвистические методы часто комбинируются. Здесь можно выделить систему ACABIT, а также TRUCKS [7].

А построение онтологии включает выделение множеств понятий и имен отношений, а также экземпляров этих понятий и связей между ними. Заполнение онтологии подразумевает наличие понятий, связей и имеет целью поиск экземпляров понятий и отношений между ними.

Следует отметить, что задачи построения и заполнения онтологий в настоящее время являются очень актуальными по следующим причинам:

1. Построение и заполнение онтологий требует разработки алгоритмов автоматического выделения информации из текстов на естественном языке. Большая часть информации в Интернет содержится именно в таком виде. Ручная обработка таких данных требует очень больших постоянно увеличивающихся человеческих ресурсов в связи с колоссальными объемами накапливаемой информации. Именно поэтому интеллектуальные алгоритмы, которые автоматизируют этот процесс, приобретают огромную важность;

2. Заполненная онтология является готовым информационным ресурсом для семантической паутины. Для реализации видения семантической паутины необходимы автоматические средства генерации метаданных. Семантическая аннотация позволяет в дальнейшем обрабатывать эту информацию машинами, воплощая концепцию Семантической паутины;

3. Заполнение онтологии может использоваться для повышения ее качества. Основная идея состоит в том, что в случае, если онтология помогает эффективно выделять необходимую информацию из текстов, можно сделать вывод, что эта онтология адекватно описывает область знания.

Построение и заполнение онтологий тесно связано с выделением информации с использованием онтологий. Основной проблемный вопрос в области выделения информации состоит в том, что-либо существующие системы нуждаются в обучении по документам, размеченным экспертами вручную, либо позволяют выделять данные только из структурированных текстов [10,11,13].

Алгоритмы, которые нуждаются в обучении, обычно эффективно работают только с документами, похожими на обучающие по стилю и жанру. В онтологии должно содержаться как можно больше информации об области знания, в частности, не только иерархия понятий и направлений, но и неиерархические (ассоциативные) связи, например, сведения о том, что конкретный метод используется в некотором направлении исследований. Требования к остальным этапам реализации алгоритма построения онтологии и ко всему алгоритму в целом:

1. Алгоритм должен извлекать иерархические и ассоциативные отношения между терминами;
2. Онтология должна отражать актуальное состояние заданной области научного знания;
3. Точность и полнота извлекаемых терминов и отношений не должна уступать показателям существующих и прошедших апробацию алгоритмов построения онтологии;
4. Алгоритм не должен требовать обучения, либо должна существовать возможность получать необходимые обучающие выборки из открытых источников, не затрачивая больших усилий по их обработке;
5. Алгоритм не должен требовать большого объема ручного труда экспертов по настройке на заданную предметную область;
6. Источники данных для алгоритма должны находиться в открытом доступе и регулярно обновляться;
7. Алгоритм должен иметь модульную архитектуру;
8. Возможность автоматизированной настройки алгоритма на конкретную область знания.

Показатели эффективности алгоритмов выделения информации делятся на два класса, а именно – показатели корректности, например, точность (корректность выделенной информации), полнота (количество выделенной информации по отношению к объему всей доступной

информации) и мера избыточности, а также оценки вычислительных ресурсов, таких как время и память.

Онтологический подход к представлению знаний позволяет применять существующие и прошедшие апробацию алгоритмы выполнения аналитических запросов. Выполнение аналитических запросов к данным обеспечивается в процессе взаимодействия конечного пользователя системы с программной реализацией модели, описывающей область знания. Запрос при использовании онтологий может выполняться автоматически с помощью механизмов логического вывода. В качестве языка запросов к онтологиям можно использовать язык SPARQL. Выбор именно этого языка обусловлен высоким уровнем его развития, зрелости и хорошим потенциалом, что подтверждают следующие факты:

- в 2008 году язык SPARQL получил статус официальной рекомендации консорциума W3C2;
- язык SPARQL не привязан к конкретному программному комплексу, в отличие от других языков запросов к онтологиям;
- для языка SPARQL существует большое число программных реализаций и приложений.

Ниже приведены примеры использования языка SPARQL:

Пример 1. Перечень направлений, которые активно исследуются в рамках интересующей области знания. Интерпретация запроса: «выдать все термины, сопоставленные результатам ... деятельности за последний (2015) год и отсортировать их по убыванию встречаемости в этих результатах». Формализуем его на языке SPARQL. Сначала сформируем множество **Terms**, содержащее все термины (возможно, с повторениями), сопоставленные результатам деятельности за последний год.

```
SELECT ?term WHERE {
  ?term a cs:term .
  ?res a swrc:Result .
  ?res swrc:isAbout ?term . ?res swrc:year 2015 .
}
```

Полученное множество терминов **Terms**, необходимо отсортировать по убыванию количества повторений каждого уникального элемента. Термины, расположенные в начале отсортированного списка, и определяют направления, которые активно исследуются в рамках интересующей области знания.

Пример 2. Перечень направлений интересов пользователя. Интерпретация запроса: «выдать список всех терминов, связанных с результатами деятельности, в авторах которых есть заданный пользователь User X». Далее представлена формальная запись этого запроса на языке SPARQL.

```
SELECT DISTINCT ?term WHERE { ?term a cs:term .
  ?res a swrc:Result .
  ?res swrc:isAbout ?term. ?res dc:creator User X .
}
```

Пример 3. Перечень форумов, посвященных интересующему направлению.

Перепишем этот запрос следующим образом: «выдать список форумов, связанных с терминами заданного направления $T = \{t_1, \dots, t_n\}$ ». Интерпретация запроса на языке SPARQL:

```
SELECT DISTINCT ?forum
WHERE {
  ? forum a swrc: Forums .
  { ? forum swrc:isAbout t_1 }
  UNION { ?forum swrc:isAbout t_2 } ...
  UNION { ?forum swrc:isAbout t_n } . }
```

Связь между запросами, формальной моделью разрабатываемой системы и кодом запросов на языке SPARQL позволяет контролировать влияние:

- модификаций множества принятых в системе запросов и используемых онтологий на программный код системы;
- модификаций программного кода системы на используемые онтологии и рассматриваемые запросы.

Последнее обстоятельство создает дополнительные возможности для эффективной верификации ПО на всех этапах её жизненного цикла [7].

В таких случаях для идентификации пользователя возможно применять методы математической лингвистики. При этом возникает необходимость в разработке:

- модели идентификатора пользователя портала сети Интернет на основе кортежа лингвистических признаков короткого сообщения;

- метода создания компонентного профиля пользователя портала сети Интернет, базирующегося на модели идентификатора, содержащего кортеж лингвистических признаков;
- методики идентификации пользователя портала сети Интернет на основе компонентного профиля.

Метод создания компонентного профиля пользователя сети Интернет предполагает реализацию ряда шагов:

- обработка пользовательских сообщений в рамках Интернет портала;
- разбор сообщений по частям речи с последующим применением шаблонов (синтаксических паттернов) для выделения наиболее распространенных конструкций;
- лексикографический анализ сообщения и выделение конструкций в соответствии с описанными шаблонами и сбор статистики об использовании знаков препинания и специальных символов;
- выделение лексических конструкций на основе слов и словоформ языка, а также выявление тематических специальных слов и выражений, характерных для аудитории конкретного форума.

На каждом шаге используются специализированные словарные базы данных и анализаторы текстовой информации, описанные в предыдущих разделах.

Заключение

Направление извлечения информации с помощью онтологий выделилось из общей задачи выделения информации сравнительно недавно, но уже отмечено как перспективное направление развития систем выделения информации. Использование онтологий является естественным развитием направления. Выделение информации традиционно имеет целью найти сведения, которые описывают некоторую область знаний, заданную структурой данных.

Онтологии же как раз и представляют собой формальную модель предметной области, выраженную, например, в виде графа понятий и связей, что обобщает иерархическую структуру данных, обычно используемую для заполнения в задаче выделения информации.

На основе исследования предметной области могут быть построены математические модели и алгоритмы, разработаны опирающиеся на онтологии архитектурные и технологические решения для создания системы пополнения и хранения, анализа и выдачи по запросу информации, характеризующей результаты деятельности *Пользователя* информации на Web-страницах в Интернет. В результате анализа состояния предметной области идентификации *Пользователей* на Web-страницах в Интернет необходимо выделить следующее:

- в связи с широкими возможностями по обеспечению анонимности пользователей порталов сети Интернет, особую важность приобретают методы идентификации. Однако, применение многих из них затруднено в связи с возможностью изменения технических характеристик устройства;
- методы определения авторства текста, применяемые классическими лингвистами, показывают хорошие результаты для больших объемов текста, подвергшихся коррекции, но требуют существенной адаптации для обработки коротких сообщений;
- в целях повышения качественных показателей методов идентификации пользователей порталов сети Интернет необходимо разработать кортеж лингвистических признаков короткого сообщения, позволяющего учитывать особенности построения идентификаторов.

С использованием онтологий и языка SPARQL возможно формальное описание запросов к системе, создающее гарантии их вычисления и дополнительные возможности для эффективной верификации кода системы на всех этапах ее жизненного цикла. А также позволяет использовать созданный прототип ПО для учета и анализа информации при моделировании и распознавании профиля *Пользователя* в социальных сетях в условиях информационной войны.

Литература

1. Jowett G. S., O'Donnel V. Propaganda and Persuasion. — Newbury Park, 1992
2. Кучумов Д.О. Проблема региональной информационной безопасности (на примере освещения в региональных СМИ террористического акта в г. Беслан) // Безопасность Евразии. – Москва, 2007, № 4. – С.53-56.
3. Russell, Stuart J. Artificial Intelligence: A Modern Approach / Stuart J Russell, Peter Norvig. — 2 edition. — Pearson Education, 2003.
4. Cowie, Jim. Information Extraction / Jim Cowie, Yorick Wilks // A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text / Ed. by Robert Dale, Hermann Moissland, Harold Somers. — New York, USA: Marcel Dekker, New York, 2000.
5. Колин К.К. Философия информации: структура реальности и феномен информации // Метафизика, 2013, № 4, 61-

84с.

6. Урсул А.Д. Природа информации: философский очерк. – 2-е изд. – Челябинск, 2010, 231 с
7. Горбунов-Посадов, М.М. Интернет-активность как обязанность ученого / М.М. Горбунов-Посадов // Информационные технологии и вычислительные системы. — 2007, № 3, 88–93с.
8. Сухопаров М.Е. Методика идентификации пользователей порталов сети интернет на основе методов математической лингвистики // 05.13.19 – Методы и системы защиты информации, информационная безопасность - Диссертация на соискание ученой степени кандидата технических наук.-Санкт-Петербург – 2015
9. Afonin S. Minimal Union-Free Decompositions of Regular Languages // Language and Automata Theory and Applications / Ed. by Adrian Dediú, Armand Ionescu, Carlos Mart'ın-Vide. — Berlin / Heidelberg: Springer-Verlag, 2009, Vol. 5457 of Lecture Notes in Computer Science, 83–92 pp.
10. Голомазов, Д.Д. Выделение терминов из коллекции текстов с заданным тематическим делением / Д.Д. Голомазов // Информационные технологии. — 2010. — № 2. — С. 8–13.
11. Васенин, В.А. и др. Использование семантических технологий для обнаружения грид-ресурсов // Программная инженерия. — 2011, № 7, 2–8 с.
12. Yeong, W. Lightweight Directory Access Protocol // Search. — 1995, Vol. 2251, №1777, 1–11pp.
13. Боровикова, О.И. Онтологический подход к построению систем информационной поддержки научной и производственной деятельности // Материалы Всероссийской конференции с международным участием «Знания – Онтологии – Теории» (ЗОНТ–09). — Т. 2. — Новосибирск: Институт математики им. С.Л. Соболева СО РАН, 2009, 93–102 с.
14. Загоруйко, Ю.А. Портал знаний по компьютерной лингвистике: содержательный доступ к лингвистическим информационным ресурсам // Компьютерная лингвистика и интеллектуальные технологии. Электронные публикации Международной конференции «Диалог-2008»
15. Aktaeva A., etc Development of a mathematical model of information warfare // International Journal of Open Information Technologies, vol. 2, №11, 2014, 28-33 pp., www.injoit.org, <https://doaj.org/toc/2307-8162>

References

1. Jowett G. S., O'Donnel V. Propaganda and Persuasion. — Newbury Park, 1992
2. Kuchumov D.O. Problema regional'noj informacionnoj bezopasnosti (na primere osvshhenija v regional'nyh SMI terroristicheskogo akta v g. Beslan) // Bezopasnost' Evrazii. – Moskva, 2007, № 4. – S.53-56.
3. Russell, Stuart J. Artificial Intelligence: A Modern Approach / Stuart J Russell, Peter Norvig. — 2 edition. — Pearson Education, 2003.
4. Cowie, Jim. Information Extraction / Jim Cowie, Yorick Wilks // A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text / Ed. by Robert Dale, Hermann Moissland, Harold Somers. — New York, USA: Marcel Dekker, New York, 2000.
5. Kolin K.K. Philosophy of information: the structure of reality and the phenomenon of information // Metafizika. 2013. № 4, 61–84 pp.
6. Ursul A.D. The nature of information: philosophical essay. Chelyabinsk, 2010. – 231 p.
7. Gorbunov-Posadov, M.M. Internet-aktivnost' kak objazannost' uchenogo / M.M. Gorbunov-Posadov // Informacionnye tehnologii i vychislitel'nye sistemy. — 2007, № 3, 88–93s.
8. Suhoparov M.E. Metodika identifikacii pol'zovatelej portalov seti internet na osnove metodov matematicheskoj lingvistiki // 05.13.19 – Metody i sistemy zashhity informacii, informacionnaja bezopasnost' - Dissertacija na soiskanie uchenoj stepeni kandidata tehniceskikh nauk.-Sankt-Peterburg – 2015
9. Afonin S. Minimal Union-Free Decompositions of Regular Languages // Language and Automata Theory and Applications / Ed. by Adrian Dediú, Armand Ionescu, Carlos Mart'ın-Vide. — Berlin / Heidelberg: Springer-Verlag, 2009, Vol. 5457 of Lecture Notes in Computer Science, 83–92 pp.
10. Golomazov, D.D. Vydelenie terminov iz kollekcii tekstov s zadannym tematiceskim de-leniem / D.D. Golomazov // Informacionnye tehnologii. — 2010. — № 2. — S. 8–13.
11. Vasenin, V.A. i dr. Ispol'zovanie semanticheskikh tehnologij dlja obnaruzhenija grid-resource // Programmnaia inzhenerija. — 2011, № 7, 2–8 pp.
12. Yeong, W. Lightweight Directory Access Protocol // Search. — 1995, Vol. 2251, №1777, 1–11pp.
13. Borovikova, O.I. Ontologicheskij podhod k postroeniju sistem informacionnoj pod-derzhki nauchnoj i proizvodstvennoj dejatel'nosti // Materialy Vserossijskoj konferencii s mezhdunarodnym uchastiem «Znanija – Ontologii – Teorii» (ZONT–09). — Т. 2. — Новосибирск: Институт математики им. С.Л. Соболева СО РАН, 2009, 93–102 pp.
14. Zagorul'ko, Ju.A. Portal znanij po komp'juternoj lingvistike: soderzhatel'nyj dostup k lingvisticheskim informacionnym resursam // Komp'juternaja lingvistika i intellektual'nye tehnologii. Jelek-tronnye publikacii Mezhdunarodnoj konferencii «Dialog-2008»
15. Aktaeva A., etc Development of a mathematical model of information warfare // International Journal of Open Information Technologies, vol. 2, №11, 2014, 28-33 pp., www.injoit.org, <https://doaj.org/toc/2307-8162>.

Поступила: 10.10.2016

Об авторах:

Актаева Алкена Умирбековна, доцент кафедры «Информационные технологии» Казахской академии транспорта и коммуникации им. М.Тынышбаева, dr.PhD, aaktaewa@list.ru;

Галиева Надежда Геннадьевна, исследователь НИЦ Павлодарского государственного университета им. С.Торыайгырова, MSc, nggaliyeva@gmail.com;

Асанова Карлыгаш, доцент Казахско-американский университет;

Наралиев Нишон Али, исследователь НИЦ Павлодарского государственного университета им. С.Торыайгырова, MSc, gbaiman@mail.ru;

Сундетов Самат, магистрант Казахской академии транспорта и коммуникации им. М.Тынышбаева, ssundetov@mail.ru;

Макулбек Назерке, магистрант Казахской академии транспорта и коммуникации им. М.Тынышбаева, nmakulbek@mail.ru.