Для цитирования: Юркова О.Н. Применение методов анализа данных для автоматизации формирования онтологии. Вестник Дагестанского государственного технического университета. Технические науки. 2018; 45 (1): 172-180. DOI:10.21822/2073-6185-2018-45-1-172-180

For citation: Yurkova O.N. Application of data analysis methods for automation of ontology formation. Herald of Daghestan State Technical University. Technical Sciences. 2018; 45 (1): 172-180. (In Russ.) DOI:10.21822/2073-6185-2018-45-1-172-180

ТЕХНИЧЕСКИЕ НАУКИ ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

УДК: 004

DOI: 10.21822/2073-6185-2018-45-1-172-180

ПРИМЕНЕНИЕ МЕТОДОВ АНАЛИЗА ДАННЫХ ДЛЯ АВТОМАТИЗАЦИИ ФОРМИРОВАНИЯ ОНТОЛОГИИ

Юркова О.Н.

Брянский государственный инженерно-технологический университет,
241037 Брянская обл., г. Брянск, пр. Станке Димитрова, 3, Россия,
e-mail: yurkova_olga@mail.ru

Резюме: *Цель.* Целью данной работы является разработка методов автоматизированного анализа текста и извлечения из полнотекстовых документов релевантных данных, а также применение методов семантического анализа текста – использованию лингвистических онтологий как формализованных моделей представления предметной области. Использование в качестве основы для построения лингвистических онтологий электронных энциклопедий, в первую очередь Википедии, с тем, чтобы извлечь из них максимум семантической информации о понятиях, их словарном выражении, взаимосвязях, а также иерархии.

Метод. Предпосылкой появления новейших технологий, которые позволили бы ее решить - это поиск решений на основе методов системного анализа, как самого текста, так и объекта исследования, которые должны быть решены в результате такой обработки. При создании современных систем искусственного интеллекта или их компонентов, разработчики и исследователи зачастую сталкиваются с необходимостью формализации определенной предметной области с целью автоматизации обработки фраз, словосочетаний и предложений, поступающих в систему на естественном языке. В настоящее время наиболее популярный подход к формальному описанию предметной области заключается в построении её онтологии. **Результат.** Описаны подходы по извлечению информации, представлена архитектура автоматизированной системы, а также приведены результаты ее применения. **Вывод.** Применены методы семантического анализа данных, использованы лингвистические онтологии как формализованные модели представления предметной области. В работе описаны подходы по извлечению информации из Википедии, представлена архитектура автоматизированной системы, а также приведены результаты ее применения.

Ключевые слова: автоматизированная система; онтология; предметная область; автоматизированное построение онтологии; Википедия; семантика

TECHNICAL SCIENCE
COMPUTER SCIENCE, COMPUTER ENGINEERING AND MANAGEMENT

APPLICATION OF DATA ANALYSIS METHODS FOR AUTOMATION
OF ONTOLOGY FORMATION

Olga N. Yurkova

Bryansk State Engineering and Technology University,
3 Stanke Dimitrova Ave., Bryansk 241037, Russia,
e-mail: yurkova_olga@mail.ru

Abstract Objectives. The aim of this work is to develop methods for automated text analysis and the retrieval of relevant data from full-text documents, as well as applying semantic text analysis methods for using linguistic ontologies as formalised models of subject area representation. Another aim is the use of electronic encyclopedias, primarily Wikipedia, as the basis for constructing the linguistic ontologies in order to derive maximum semantic information about their concepts, vocabulary expressions, interrelations and hierarchy. **Methods.** The search for solutions based on system analysis methods is based on the emergence of new technologies that for solving both the text itself and the object of research that is to be solved as a result of such processing. When creating contemporary artificial intelligence systems or their components, developers and researchers often face the need to formalise a certain subject area in order to automate the processing of phrases, word collocations and sentences entering the system in natural language form. Currently, the most popular approach to the formal description of a subject area is to construct an ontology. **Results.** Established approaches to the retrieval of information are described along with the architecture of the automated system and the results of their application. **Conclusion.** Semantic data analysis methods are applied with linguistic ontologies used as the formalised models of subject area representation. Approaches to retrieving information from Wikipedia are described along with the architecture of the automated system and results of its application.

Keywords: automated system, ontology, subject area, automated ontology construction, wikipedia, semantics

Введение. В настоящее время в связи с увеличением потока информации появилась необходимость поиска новых способов ее хранения, представления, формализации и систематизации, а также автоматической обработки.

Растет интерес к всеобъемлющим базам знаний, которые возможно использовать для различных практических целей. Как результат, на фоне вновь возникающих потребностей развиваются новые технологии, призванные решить заявленные проблемы.

Особую важность приобретает обеспечение эффективного использования данных с применением интеллектуальных средств их анализа и представления. Полезность знаний в выборе того или иного решения выше в том случае, когда эти знания систематизированы и формализованы.

При создании современных систем искусственного интеллекта или их компонентов разработчики и исследователи зачастую сталкиваются с необходимостью формализации определенной предметной области с целью автоматизации обработки фраз, словосочетаний и предложений, поступающих в систему на естественном языке. В настоящее время наиболее популярный подход к формальному описанию предметной области заключается в построении её онтологии.

Онтология представляет собой структурное описание предметной области, включающее словари, термины, отношения. Основные элементы (примитивы) онтологии – это персоны (примеры), типы данных (конкретные домены), аксиомы, концепции (понятия, классы). Обобщая множество сходных определений, с уверенностью можно сказать, что в настоящее время

под онтологией подразумевается семантическая структура, описывающая основные концепты предметной области, связывающие их отношения, а также аксиомы, описывающие правила вывода новых концептов [1].

Более подробно понятие онтологии сформулировано у Т.А. Гавриловой, и описано как точная спецификация некоторой предметной области, представляющая собой формальное и декларативное описание, которое включает словарь терминов предметной области и логические выражения, описывающие значения этих терминов, а также то, как они соотносятся между собой [2]. До недавних пор для решения задач анализа текста, в большинстве случаев, применялись и развивались статистические методы.

С появлением сетевых лексических ресурсов и с совершенствованием синтаксических анализаторов лингвистические методы привлекают все большее внимание специалистов благодаря высокому качеству результатов, которые обещают такие методы [3-5]. Поэтому целью данной работы является описание структуры ИС, позволяющей обнаруживать в текстовом корпусе релевантные концепты предметной области на основе использования лингвистических шаблонов, характерных для данной предметной области.

Основной принцип построения онтологий можно разбить на несколько независимых этапов, на каждом из которых решается одна определенная задача, результаты которой, в свою очередь, служат исходными данными для задачи следующего, как правило, более сложного уровня.

Опираясь на этот подход, можно выделить некоторую последовательность действий: извлечение из текста разбиения терминов на → термины-кандидаты присвоения → группы (кластеризация) → обобщающего концепта каждой группе → определение отношений между концептами расширения концептов. Для начальных этапов многочисленные статистические методы позволяют получать весьма качественные результаты (до 90% точности по сравнению с результатами работы экспертов [7]). Однако для построения таксономии понятий и обнаружения отношений необходима разработка лингвистических методов. В частности, одним из эффективных подходов для обнаружения группы синонимов и общего для них гиперонима является использование лингвистических шаблонов, предложенных еще М. Hearst [8]. Совершенствование данного подхода зависит от наличия общей теории описания и анализа конкретного языка.

В данном случае, автор разделяет идею Золотовой Г. А. о том, что синтаксический строй текста организуется «элементарными» единицами. В качестве такой единицы выдвинуто понятие синтаксемы. Синтаксемой по Золотовой Г.А. называется минимальная, неделимая семантико-синтаксическая единица русского языка, выступающая одновременно и как носитель элементарного смысла, и как конструктивный компонент с функциональностью, необходимой и достаточной для построения более сложных синтаксических конструкций.

Формально онтология описывается следующим образом [11]:

$$O = \{C, R, A\},$$

где O – онтология, C – совокупность понятий предметной области, R – совокупность отношений между ними, A – набор аксиом.

Онтология обеспечивает словари для представления и обмена знаниями о некоторой предметной области и множество связей, установленных между терминами в этих словарях.

В основе онтологического анализа лежит описание системы в терминах сущностей, отношений между ними и преобразование сущностей, которое выполняется в процессе решения определенной задачи. Онтологический инжиниринг подразумевает глубокий структурный анализ предметной области. Основным преимуществом онтологического инжиниринга является целостный подход к автоматизации предприятия.

Формализованное представление понятийной структуры предметной области в онтологии делает возможной автоматическую обработку онтологической информации, что находит активное применение в современных поисковых системах [12], мультиагентных системах [13], системах интеграции данных, получаемых из гетерогенных источников [14] и в других задачах.

Онтологии предметных областей применяются в таких областях, как:

1. Построение семантически ориентированных поисковых систем и систем обра-

ботки естественно-языковых текстов. Примерами являются проект Exactus [15] и система Texterra [16].

2. Разработка интеллектуальных и экспертных систем [17].
3. Разработка автоматизированных систем мониторинга данных для социально-экономических систем [18].
4. Управление знаниями [9].
5. Построение специализированных научных информационных систем, позволяющих унифицировать терминологию, используемую специалистами в рамках определенной предметной области [20].
6. Системный анализ предметной области.
7. Создание электронных обучающих систем, использующих онтологическое структурирование элементов знаний и учебных объектов [14].

Среди основных современных направлений, в которых ведется исследовательская работа по созданию средств и технологий автоматизации формирования онтологий предметных областей, можно выделить следующие.

1. Автоматизированное построение онтологии на основе анализа и мониторинга содержания специальных web-ресурсов. В первую очередь речь идет о ресурсах, разработанных в рамках проекта SymanticWeb, которые включают описанные в XML-подобных форматах (XML, XML Schema, RDF, RDF Schema, OWL, HTMLmicrodata) метаданные, содержащие дополнительные сведения о взаимосвязи используемых на ресурсах понятий. Используя такую семантическую разметку web-страниц, можно автоматизировать построение соответствующих онтологий.
2. Автоматизированный синтез онтологии на основе анализа массива текстов предметной области. В этом подходе процесс построения онтологии разбит на несколько независимых этапов, на каждом из которых решается одна определенная задача, результаты которой, в свою очередь, служат исходными данными для задачи следующего, как правило, более сложного уровня [18].
3. Автоматизированное выделение онтологии предметной области из универсальных онтологий (например, электронных энциклопедий). Так, в работе [19] описан положительный опыт по автоматизации построения онтологии предметной области на основе структуры категорий Википедии, а сам подход на данный момент является перспективным и востребованным.

То, чего действительно не хватает для полноценной интеллектуальной системы, так это систематизированного репертуара лексико-синтаксических единиц языка, несущих в себе однозначно трактуемую семантику и одновременно выполняющих роль «элементарных единиц сборки» высказываний (текстов). Пересечение такого репертуара «архе-функций», несомых синтаксической формой и репертуара слов-носителей категориальных референций, дает проекцию однозначно трактуемой роли выполняемой данным понятием. Подобный систематизирующий труд, как было сказано выше, выполнен в значительной степени Золотовой Г.А. [8] и ждет достойной технической реализации.

Постановка задачи. Наиболее перспективным направлением автору видится разработка алгоритмов и программной системы, которая бы позволяла строить лингвистические онтологии заданной предметной области на основе анализа и мониторинга данных, получаемых из различных источников. На начальном этапе разработки такой системы в качестве источника данных можно ограничиться Википедией, а после успешного апробирования алгоритмов на ней можно постепенно дополнять список источников.

Методы исследования. Для обработки страниц Википедии потребуется разработка специального модуля – парсеравики-страниц для извлечения необходимой информации. Причем данный модуль будет ориентирован на анализ русскоязычной версии Википедии, и его разработка является достаточно трудоемкой задачей. Поэтому на этапе исследования в целях проверки возможности эффективного извлечения онтологии заданной предметной области, разумно воспользоваться уже имеющимися общедоступными хранилищами извлеченной из Вики-

педии, например DBpedia, YAGO или Wikidata. В качестве источника для импорта данных в рамках собственного проекта было решено использовать именно DBpedia.

В основе Википедии лежит так называемый Вики-движок – комплекс программных средств для преобразования Вики-разметки в код, предназначенный для отображения в браузере. Одним из самых распространенных движков является MediaWiki, на его основе и работает Википедия. Для него существуют специальные дополнения, называемые расширениями, позволяющие получить определенную функциональность.

Так, расширение Semantic MediaWiki дает возможность добавлять семантическую информацию за счет расширения разметки, а также предлагает средства для работы с этой информацией. Для проведения работ технического характера в Wiki-системах используют боты – специальные программы для выполнения заданного набора операций. Они являются клиентскими приложениями, поэтому не требуют внесения изменений на стороне сервера. В Википедии боты используются для таких задач, как переименование категорий и статей, расстановка интервики-ссылок, исправление ссылок, удаление спама и т.п.

Для реализации ботов используются различные языки программирования, а также существуют разные библиотеки для облегчения их написания. Одной из наиболее развитых библиотек является Python WikipediaBot Framework. Она использует MediaWiki API (специальный интерфейс прикладного программирования) для взаимодействия с MediaWiki-системой – авторизации, получения данных и внесения изменений.

Онтология из Википедии извлекается в следующем порядке. Сначала извлекаются все классы, при этом каждому классу соответствует одна категория Wiki, а структура вложенности категорий Wiki определяет иерархию классов. Затем извлекаются все страницы как экземпляры соответствующих классов. Для пустых страниц, на которые в Wiki имеются ссылки, заводится специальный служебный класс «Несуществующие страницы». После этого просматриваются все ссылки на каждой странице. Для начала определяется, является ли ссылка обычной или семантической. Если ссылка обычная, то для соответствующего экземпляра класса в OWL-онтологии заводится объектное свойство «Ссылается на» (так как ссылка обычная, а не семантическая, то у нее нет своего собственного имени, и данное имя выбрано для всех таких ссылок) со значением в виде экземпляра, имя которого совпадает с именем страницы, на которую указывает ссылка. Если ссылка семантическая, то она имеет структуру, и для нее сначала определяется тип ее свойства. Если свойство имеет тип «Страница» или его тип не указан, то в OWL-онтологии заводится объектное свойство с соответствующими именем (название свойства) и значением (значение свойства). Следует заметить, что по умолчанию свойство ссылки имеет тип «Страница».

Следует отметить, что Wiki-система, из которой извлекается онтология, не обязательно должна функционировать с расширением Semantic MediaWiki. Однако в случае использования Википедии без этого расширения извлекаемая онтология будет гораздо беднее, так как в ней не будет присутствовать специальная семантическая информация. В частности, нельзя будет извлечь атрибуты, разнообразие отношений также будет невелико. Правда, путем индивидуальной настройки на конкретную Wiki-систему объем извлекаемой из нее информации можно увеличить.

В рамках работы основной акцент сделаем на обработку следующих наборов данных.

1. Ссылки на страницах статей (файл `page_links_ru`). К ним относятся гиперссылки, содержащиеся непосредственно в текстах статей. Данный набор данных самый большой (в ходе разбора этого набора данных было выделено порядка 33,68 млн. ссылок и 5,39 млн. понятий для русскоязычной DBpedia).
2. Ссылки на категории, к которым относится понятие, описываемое в статье (файлы `article_categories_ru` и `skos_categories_ru`). Для большинства статей Википедия в специальном блоке перечисляются категории, к которой относится описываемое понятие. При этом категории могут содержать подкатегории и служат инструментом для объединения статей одной тематики. Однако структура категорий Википедии сетевая, а не строго иерархическая, что затрудняет классификацию понятий,

а также создает дополнительные трудности для выделения онтологии предметной области, т.к. нельзя ограничиваться только категориями статьи [20].

3. Ссылки на дополнительные свойства понятия, описываемого в статье, в специальных инфобокс-секциях (файл infobox_properties_ru). Обычно эта секция располагается в правом верхнем углу статьи и в ней содержатся основные факты или статистические данные о предмете статьи. При этом зачастую статьи со схожей проблематикой содержат в этом поле одинаковые свойства [20].
4. Страницы-перенаправления (файл redirects_ru). Данный набор данных можно рассматривать как основу для определения синонимов.

Итак, существуют подходы для извлечения таксономии из MediaWiki-систем без семантических расширений (например, в проектах YAGO и DBpedia, упомянутых выше), есть стандартные средства для экспорта семантической информации в расширение Semantic MediaWiki, но нет автоматизированных программных систем, объединяющих в себе и то, и другое. И именно в этом заключается новизна данной работы.

Общая архитектура разрабатываемой автоматизированной программной системы, решающей задачу извлечения онтологий предметных областей из внешних источников, представлена на рис. 1.

Рис. 1. Архитектура программной системы автоматизированного построения онтологий предметных областей

Fig. 1. The architecture of the software system for the automated construction of ontologies of subject domains

В программной системе автор выделяет 3 модуля и 3 вида баз данных.

1. Подсистема импорта предназначена для импорта данных из внешних источников знаний, представленных в электронном виде, во внутренние базы данных программной системы для дальнейшей обработки.
2. База данных понятий и связей содержит основную информацию о списке понятий и связях между ними (с указанием типа связи), полученных в ходе импорта.
3. Подсистема построения онтологии предназначена для извлечения из импортированных данных понятий предметной области и построения на их основе лингвистической онтологии предметной области с соответствующей внутренней структурой.
4. Подсистема взаимодействия с пользователем-экспертом предназначена для интерактивного взаимодействия с экспертами на этапах извлечения онтологии соответствующей предметной области.

5. Хранилище универсальной онтологии содержит онтологию, полученную на всем массиве входных данных, включающую множество предметных областей. Для оптимизации общей производительности оно частично заполняется уже на этапе импорта. В дальнейшем его дополняет подсистема построения онтологии.
6. Хранилища онтологий предметных областей предназначены для сохранения онтологий конкретных предметных областей, формируются подсистемой построения онтологии. Они намного меньше по объему, чем хранилище универсальной онтологии, и могут быть использованы для эффективной обработки запросов к онтологии. Структура этого хранилища соответствует структуре хранилища универсальной онтологии.

Обсуждение результатов. Разработанная программная система была проверена на предметной области «Информационные технологии». В качестве исходных понятий были взяты следующие наиболее общие понятия, соответствующие одноименным категориям Википедии предметной области информационных технологий: «Информационные технологии», «Программное обеспечение», «Компьютерные науки» и «Информатика».

В ходе исследования для разных вариантов пороговых значений были получены соответствующие варианты онтологий предметной области. Результаты исследований представлены в табл. 1.

Степень соответствия полученной онтологии предметной области и множества понятий, включенных в него, оценивалась экспертным путем авторами работы и градуировалась по 10-балльной шкале (1 – минимальное соответствие, 10 – максимальное соответствие).

В ходе этой процедуры эксперту случайным образом выдавалось 100 понятий, и для каждого из них он оценивал, соотносится ли оно с предметной областью или нет. Чем больше понятий соответствовало предметной области, тем выше степень соответствия.

Таблица 1. Результаты работы разработанного алгоритма по выделению предметной области «Информационные технологии»

Table 1. The results of the developed algorithm for the selection of the subject area «Information Technology»

№ эксперимента Experiment No.	Количество понятий в предметной области The number of concepts in domain	Степень соответствия Degree of conformity
1	82053	3
2	52780	4
3	20822	6
4	125732	1
5	91282	2
6	32445	4
7	75878	3
8	43370	5
9	14708	8
10	110752	1
11	65590	3
12	18666	6

Вывод. По результатам экспериментов можно сделать вывод, что автоматизированная система по автоматическому извлечению онтологии предметной области на основе данных, полученных из DBPedia, показала свою работоспособность.

Предлагаемый подход автоматизированного построения онтологии предметной области сочетает в себе быстроту статистических методов и точность лингвистического подхода с позиций синтаксиса (или лингвистических шаблонов).

Процесс построения онтологии состоит из нескольких последовательных этапов.

В данной работе предложен подход к автоматизированному построению (расширению) базовой онтологии на основе синтаксиса русского языка. Поскольку каждая синтаксема описывается конечным детерминированным множеством признаков, такой подход является не только возможным, но и предпочтительным, поскольку он обеспечивает однозначное определение свойств концептов создаваемой онтологии.

Библиографический список:

1. Аверченков В.И. Мониторинг и системный анализ информации в сети Интернет: монография / Аверченков, В.И., Рошин, С.М. //Брянск: БГТУ, 2006. - 160 с.
2. Барсегян А.А. Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP / Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. // СПб.: БХВ-Петербург, 2007. – 384 с.: ил.
3. Гаврилова Т. А. Онтологический подход к управлению знаниями при разработке корпоративных информационных систем / Т.А. Гаврилова // «Новости искусственного интеллекта». 2003. №2. С. 24-30
4. Копелиович Д.И. Принципы построения автоматизированных систем мониторинга социально-экономических объектов / Копелиович Д.И., Юркова О.Н.//Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика, 2015- № 1 - С. 98-104.
5. Копелиович Д.И. Функциональное моделирование процесса мониторинга данных/Копелиович Д.И., Рыженков Д.А.//Мониторинг. Наука и технологии. 2016. № 1. С. 49-53
6. Кравцов Д.В. Автоматизированная система для построения онтологий предметных областей / Кравцов Д.В., Коростелёв Д.А., Юркова О.Н.// МОНИТОРИНГ. Наука и Технологии, 2017. № 1 (30). – С. 46-50
7. Найханова Л.В. Основные аспекты построения онтологий верхнего уровня и предметной области // В сборнике научных статей “Интернет-порталы: содержание и технологии”. Выпуск 3. / Редкол.: А.Н. Тихонов (пред.); ФГУ ГНИИ ИТТ “Информатика”. – М.: Просвещение, 2005. – С. 452-479
8. Лапшин В.А. Онтологии в компьютерных системах/ В.А. Лапшин. - М.: Научный мир. 2010.- 222 с.
9. Теслинова Е.А. Разработка онтологии системы управления знаниями организации с использованием методологии концептуального проектирования / Е.А. Теслинова// Успехи современного естествознания. – 2006. – № 9. – С. 96-98
10. Евгеньев Г.Б. Интеллектуальные системы проектирования / Г.Б. Евгеньев. – М., Изд-во МГТУ им. Н.Э. Баумана, 2009. – 334 с.
11. Gruber T. A translation approach to portable ontologies / T. Gruber // Knowledge Acquisition. – 1993. – Vol. 5. – P. 199-220.
12. Studer, R. Knowledge Engineering: Principles and methods / R. Studer, R. Benjamins, D. Fensel // Data and knowledge engineering. – 1998. Vol. 25. – P. 161-197.
13. Yildiz B., Miksch S. Ontology-Driven Information Systems: Challenges and Requirements / B.Yildiz, S.Miksch // International Conference on Semantic Web and Digital Libraries. – Indian Statistical Institute Platinum Jubilee Conference Series. –2007. – P. 35-44.
14. Мирошников В.В. Онтологическая модель системы управления знаниями в области качества / В.В.Мирошников, Д.И.Булатицкий // Вестн. БГТУ. – 2009. № 4. – С. 100-106.
15. Аверченков В.И. Формализация описания предметной области «Обеспечение технологичности конструкций изделий в интегрированных САПР» на основе онтологии / В.И.Аверченков, В.А.Шкаберин // Справочник. Инженерный журнал. – Брянск. 2009. – № 10. –С. 32-38.
16. Смирнов С.В. Онтологический анализ предметных областей моделирования / С.В.Смирнов// Известия Самарского научного центра РАН. – 2001. Т. 3.№ 1.– С.62-70.
17. Антонов И.В. Формирование онтологических моделей предметной области для электронных обучающих систем / И.В.Антонов, М.В.Воронов// Информационные технологии в обеспечении нового качества высшего образования. Сборник научных статей. – Кн. 2. – М.: Исследовательский центр проблем качества подготовки специалистов. 2010. – С. 48–55.
18. Buitelaar P., Cimiano P., Magnini B. Ontology Learning from Text: Methods //Evaluation and applications. IOS Press. – 2005.
19. Korshunov, A. A Category-Driven Approach to Deriving Domain Specific Subset of Wikipedia / A.Korshunov, D.Turdakov, J.Jeong, M.Lee, C.Moon // SYRCoDIS. – 2011. –P.43-53.
20. Варламов М.И. Расчет семантической близости концептов на основе кратчайших путей в графе ссылок Википедии / М.И.Варламов, А.В.Коршунов// Машинное обучение и анализ данных. – 2014. Т.1, №8. – С. 1107–1125.
- rilova T.A. Ontological approach to knowledge management in the development of corporate information systems. Novosti iskusstvennogo intellekta. 2003;2:24-30. (In Russ.)]
4. Kopeliovich D.I., Yurkova O.N. Printsipy postroeniya avtomatizirovannykh sistem monitoringa sotsial'no-ekonomicheskikh ob"ektov. Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naya tekhnika i informatika. 2015;1:98-104. [Kopeliovich D.I., Yurkova O.N. Principles of constructing automated monitoring systems for socio-economic objects. Bulletin of the Astrakhan State Technical University. Series: Management, Computer Engineering, Computer Science. 2015;1:98-104. (In Russ.)]
5. Kopeliovich D.I., Ryzhenkov D.A. Funktsional'noe modelirovanie protsesssa monitoringa dannykh. Monitoring. Nauka i tekhnologii. 2016;1:49-53. [Kopeliovich D.I., Ryzhenkov D.A.

References

1. Averchenkov V.I., Roshchin C.M. Monitoring i sistemnyi analiz informatsii v seti Internet. Bryansk: BGTU; 2006. 160 s. [Averchenkov V.I., Roshchin C.M. Monitoring and system analysis of information on the Internet. Bryansk: BGTU; 2006. 160 p. (In Russ.)]
2. Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I. Tekhnologii analiza dannykh. Data Mining, Visual Mining, Text Mining, OLAP. SPb.: BKhV-Peterburg; 2007. 384 s. [Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I. Technologies of data analysis. Data Mining, Visual Mining, Text Mining, OLAP. SPb.: BKhV-Peterburg; 2007. 384 p. (In Russ.)]
3. Gavrilova T.A. Ontologicheskii podkhod k upravleniyu znaniyami pri razrabotke korporativnykh informatsionnykh sistem. Novosti iskusstvennogo intellekta. 2003;2:24-30. [Gav-

- Functional modeling of the data monitoring process.. Monitoring. Nauka i tekhnologii. 2016;1:49-53. (In Russ.)]
6. Kravtsov D.V., Korostelev D.A., Yurkova O.N. Avtomatizirovannaya sistema dlya postroeniya ontologii predmetnykh oblastei. Monitoring. Nauka i Tekhnologii, 2017;1(30):46-50. [Kravtsov D.V., Korostelev D.A., Yurkova O.N. Automated system for constructing ontologies of subject areas. Monitoring. Nauka i Tekhnologii, 2017;1(30):46-50. (In Russ.)]
7. Naikhanova L.V. Osnovnye aspekty postroeniya ontologii verkhnego urovnya i predmetnoi oblasti. Sbornik nauchnykh statei "Internet-portaly: soderzhanie i tekhnologii". FGU GNII ITT "Informatika". M.: Prosveshchenie. 2005;3:452-479. [Naikhanova L.V. The main aspects of constructing ontologies of the top level and subject domain. Collection of scientific articles "Internet portals: content and technology". FGU GNII ITT "Informatika". M.: Prosveshchenie. 2005;3:452-479. (In Russ.)]
8. Lapshin V.A. Ontologii v komp'yuternykh sistemakh. M.: Nauchnyi mir; 2010. 222 s. [Lapshin V.A. Ontologies in computer systems. M.: Nauchnyi mir; 2010. 222 p. (In Russ.)]
9. Teslinova E.A. Razrabotka ontologii sistemy upravleniya znaniyami organizatsii s ispol'zovaniem metodologii kontseptual'nogo proektirovaniya. Uspekhi sovremennogo estestvoznaniya. 2006;9:96-98. [Teslinova E.A. Development of the ontology of the organisation's knowledge management system using the conceptual design methodology. Advances in current natural sciences. 2006;9:96-98. (In Russ.)]
10. Evgenev G.B. Intellektual'nye sistemy proektirovaniya. M.: Izd-vo MGTU im. N.E. Bauman; 2009. 334 s. [Evgenev G.B. Intelligent design systems. M.: Izd-vo MGTU im. N.E. Bauman; 2009. 334 p. (In Russ.)]
11. Gruber T. A translation approach to portable ontologies. Knowledge Acquisition. 1993;5:199-220.
12. Studer R., Benjamins R., Fensel D. Knowledge Engineering: Principles and methods. Data and knowledge engineering. 1998;25:161-197.
13. Yildiz V., Miksch S. Ontology-Driven Information Systems: Challenges and Requirements. International Conference on Semantic Web and Digital Libraries. Indian Statistical Institute Platinum Jubilee Conference Series. 2007. P. 35-44.
14. Miroshnikov V.V., Bulatitskii D.I. Ontologicheskaya model' sistemy upravleniya znaniyami v oblasti kachestva. Vestnik BGTU. 2009;4:100-106. [Miroshnikov V.V., Bulatitskii D.I. Ontological model of the knowledge management system in the field of quality. Vestnik BGTU. 2009;4:100-106. (In Russ.)]
15. Averchenkov V.I., Shkaberin V.A. Formalizatsiya opisaniya predmetnoi oblasti "Obespechenie tekhnologichnosti konstruktivnykh izdelii v integrirovannykh SAPR" na osnove ontologii. Spravochnik. Inzhenernyi zhurnal. 2009;10:32-38. [Averchenkov V.I., Shkaberin V.A. Formalisation of the description of the "Ensuring the manufacturability of product designs in integrated CAD systems" subject area on the basis of ontology. Handbook. Engineering Journal. 2009;10:32-38. (In Russ.)]
16. Smirnov S.V. Ontologicheskii analiz predmetnykh oblastei modelirovaniya. Izvestiya Samarskogo nauchnogo tsentra RAN. 2001;3(1):62-70. [Smirnov S.V. Ontological analysis of modeling subject areas. Izvestia of Samara Scientific Center of the Russian Academy of Sciences. 2001;3(1):62-70. (In Russ.)]
17. Antonov I.V., Voronov M.V. Formirovanie ontologicheskikh modelei predmetnoi oblasti dlya elektronnykh obuchayushchikh sistem. Informatsionnye tekhnologii v obespechenii novogo kachestva vysshego obrazovaniya. Sbornik nauchnykh statei. Kn. 2. M.: Issledovatel'skii tsentr problem kachestva podgotovki spetsialistov. 2010. S. 48-55. [Antonov I.V., Voronov M.V. Formation of ontological models of the subject area for electronic learning systems. Information technology in ensuring a new quality of higher education. Collection of scientific articles. Book 2. M.: Issledovatel'skii tsentr problem kachestva podgotovki spetsialistov. 2010. S. 48-55. (In Russ.)]
18. Buitelaar P., Cimiano P., Magnini B. Ontology Learning from Text: Methods. Evaluation and applications. IOS Press. 2005.
19. Korshunov A.A., Turdakov D., Jeong J., Lee M., Moon C. Category-Driven Approach to Deriving Domain Specific Subset of Wikipedia. SYRCoDIS. 2011. P.43-53.
20. Varlamov M.I., Korshunov A.V. Raschet semanticheskoi blizosti kontseptov na osnove kratchaishikh putei v grafe ssylok Vikipedii. Mashinnoe obuchenie i analiz dannykh. 2014;1(8):1107-1125. [Varlamov M.I., Korshunov A.V. Calculation of semantic proximity of concepts based on shortest paths in the Wikipedia link graph. Machine Learning and Data Analysis. 2014;1(8):1107-1125. (In Russ.)]

Сведения об авторах:

Юркова Ольга Николаевна – кандидат экономических наук, доцент, кафедра информационных технологий.

Information about the author.

Olga N. Yurkova – Cand. Sci.(Economics), Assoc. Prof., Department of Information Technologies

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию 14.12.2017.

Принята в печать 29.01.2018.

Conflict of interest.

The author declare no conflict of interest.

Received 14.12.2017.

Accepted for publication 29.01.2018.