

19. Kureychik V.M., Kazharov A.A. Ispol'zovanie shablonnykh resheniy v murav'inykh algoritмах [The use of standard solutions in ant colony optimization algorithms], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2013, No. 7 (144), pp. 11-17.
20. Gladkov L.A., Kravchenko Y.A., Kureichik V.V. Evolutionary Algorithm for Extremal Subsets Comprehension in Graphs, *World Applied Sciences Journal*, 2013, No. 27 (9), pp. 1212-1217.

Статью рекомендовал к опубликованию д.т.н., профессор А.Н. Афанасьев.

Кравченко Юрий Алексеевич – Южный федеральный университет; e-mail: yakravchenko@sfedu.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

Курситыс Илона Олеговна – e-mail: i.kursitys@mail.ru; кафедра систем автоматизированного проектирования; аспирант.

Kravchenko Yury Alekseevich – Southern Federal University; e-mail: yakravchenko@sfedu.ru; 44, Nekrasovskiy lane, Taganrog, 347928, Russia; phone: +78634371651; the department of computer aided design; associate professor.

Kursitys Iona Olegovna – e-mail: i.kursitys@mail.ru; the department of computer aided design; postgraduate student.

УДК 002.53:004.89

А.А. Новиков, Ю.С. Старкова, В.В. Марков, Д.Ю. Кравченко

СЕМАНТИЧЕСКИЙ ПОИСК С ИСПОЛЬЗОВАНИЕМ ГЕНЕТИЧЕСКИХ ОПЕРАТОРОВ*

Традиционные механизмы поиска основываются на поиске по ключевым словам, в котором совсем не учитываются семантические связи между различными понятиями. Это приводит к потере релевантных документов из-за неточной формулировки запроса или использования в запросе близких по смыслу слов и понятий. Для решения проблем формулировки запросов пользователей и междисциплинарности понятий предлагается использовать семантический поиск. Предлагаемый в статье подход реализации семантического поиска применим к большим объёмам текстовых данных и основан на использовании генетического алгоритма. В отличие от стандартных подходов к поиску информации, описываемый подход позволяет учитывать семантику взаимосвязей между понятиями, а также корректно обрабатывать междисциплинарные понятия. Благодаря семантической индексации в документах определяются понятия, не присутствующие в исходном запросе пользователя, но являющиеся семантически близкими с понятиями из запроса. Семантическая индексация производится для каждого документа отдельно, что позволяет производить параллельную индексацию сразу по нескольким предметным областям. К моменту завершения формирования онтологического профиля рассматриваемого документа вычисляются все семантические расстояния между парами выделенных понятий. Понятия считаются близкими по смыслу, если их значение семантической близости выше некоторого порогового значения, которое задается в параметрах поиска. Построение онтологического профиля документа является многокритериальной задачей, так как зависит от множества характеристик, поэтому для ее решения можно использовать генетические алгоритмы. Разработанный генетический алгоритм предназначен для более точного распределения весовых коэффициентов и оценки семантической близости понятий.

Семантический поиск; информационный поиск; онтология; генетический алгоритм.

* Работа выполнена при финансовой поддержке РФФИ (проект № 16–07–00703).

A.A. Novikov, Yu.S. Starkova, V.V. Markov, D.Yu. Kravchenko

SEMANTIC SEARCH WITH THE USE OF GENETIC OPERATORS

Traditional search mechanisms are based on keyword search, which does not take into account the semantic links between different concepts. This leads to the loss of relevant documents due to inaccurate wording of the query or the use in the query of related words and concepts. To solve the problems of formulating user requests and interdisciplinarity of concepts, it is suggested to use the semantic search. The approach proposed in the article for the implementation of semantic search is applicable to large volumes of text data and is based on the use of a genetic algorithm. Unlike standard approaches to information retrieval, the described approach allows taking into account the semantics of interrelations between concepts, as well as correctly handle interdisciplinary concepts. Thanks to semantic indexing, documents define concepts not presented in the user's initial query, but semantically close to the concepts from the query. Semantic indexing is performed for each document separately, which allows parallel indexation on several subject areas. By the time of completion of the formation of the ontological profile of the document in question, all semantic distances between pairs of distinguished concepts are calculated. Concepts are considered close in meaning if their semantic proximity value is above a certain threshold value that is specified in the search parameters. Building an ontological document profile is a multicriteria task, since it depends on a lot of characteristics, so genetic algorithms can be used to solve it. The developed genetic algorithm is intended for more accurate distribution of weight coefficients and estimation of semantic proximity of concepts.

Semantic search; information retrieval; ontology; genetic algorithm.

Введение. Стандартные механизмы анализа текстов в большинстве своем фокусируются на статистическом анализе, не учитывающем семантические связи между понятиями. Синтаксические подходы отлично работают в задачах анализа текста, но они имеют ряд ограничений во время выполнения контекстного поиска информации. Поиск по ключевым словам, используемый во всех традиционных поисковых механизмах, может не найти необходимый пользователю документ, если запрос сформулирован с использованием близких по смыслу слов или терминов, которые отсутствуют в искомом документе.

Двумя главными проблемами здесь являются [1]:

1. Формулировка запроса. Если пользователь выражает свой запрос с использованием близких по смыслу терминов, синонимов или обобщений, то в результирующий набор не попадут многие релевантные документы.
2. Междисциплинарность. Одни и те же понятия в различных областях наук могут иметь различные значения. Например, термин «онтология» в философии и информатике имеют разные значения.

Для решения этих проблем необходимо разработать алгоритм, позволяющий учитывать знания из онтологии, которая описывает предметную область содержащихся коллекций документов, с использованием семантических метрик, которые учитывают контекст запроса, на примере анализа научных публикаций. Для корректной обработки междисциплинарных понятий и усовершенствования механизма индексации запросов используется генетический алгоритм [2, 3].

Несмотря на различные требования к оформлению, научные публикации имеют шаблонизированную структуру, которая включает в себя: заголовок, аннотация, ключевые слова, основной текст документа и заключение. В разрабатываемый алгоритм будет заложен учет данного шаблона для всех статей. Чаще всего наиболее важные понятия выносятся в заголовок, аннотацию или находятся в списке ключевых слов. Эти понятия будут оцениваться выше чем другие понятия, которые появляются лишь в основном тексте документа.

1. Семантический поиск и построение семантического профиля документа. Существуют три основных подхода к осуществлению семантического поиска: кластеризация полученных результатов, анализ синтаксиса и семантики естественного языка и поиск с учетом семантических связей между понятиями. Первые два подхода увеличивают точность поиска путем правильной семантической интерпретации понятий и разрешения конфликтов во время интерпретации междисциплинарности, но они находят лишь те документы, в которых встречается хотя бы одно ключевые слово из запроса. Последний подход предоставляет возможность разбиения понятий на триплеты («субъект, свойство, объект»). Он позволяет находить релевантные документы, которые не содержат ключевых слов из запроса пользователя, но имеют семантические связи с ними. Для описания различных понятий предметной области и указания связей между ними используются онтологии [3–5].

Чтобы улучшить качества семантического поиска необходимо использовать дополнительные источники знаний, которые будут описывать различные понятия и связи между ними. Для этого используются хранилище онтологий. Также необходим новый механизм индексации документов и запросов пользователей, который будет учитывать не только понятия, встречающиеся в документе, но и понятия, имеющие семантическую связь.

Механизм индексации документов и запросов должен учитывать статистические характеристики, такие как [6]:

- ◆ количество повторений понятия в документе;
- ◆ документная частота;
- ◆ дисперсия понятия;
- ◆ структура искомого документа.

Исходя из этого можно предположить, что имеет место задача многокритериальной оптимизации, для решения которой можно использовать генетический алгоритм (ГА). При правильной настройке учитываемых параметров ГА выдает результат, очень близкий к оптимальному за полиномиальное время. Генетические алгоритмы также хорошо справляются с решением задач оптимизации при отсутствии четких требований к целевой функции. Для индексации документов и запросов пользователя необходимо сформировать онтологический профиль документа. Онтологический профиль документа – список пар «понятие, весовой коэффициент», относящихся к определенной предметной области, куда входят только те понятия из онтологии, которые были найдены в тексте документа, либо связаны с понятиями из документа семантически, причем длина семантической связи (семантическое расстояние) задается в параметрах ГА с помощью порогового значения. Здесь семантической индексацией документов является формирование онтологического профиля с учетом метаданных имеющихся онтологий [6–10].

Свойствами онтологического профиля являются [11]:

- 1) зависимость от онтологии;
- 2) содержание понятий, состоящих из нескольких слов;
- 3) содержание всех, семантически связанных понятий из онтологии;
- 4) содержание всех понятий, удовлетворяющих установленным критериям.

Все документы индексируются независимо друг от друга. При этом индексация документов по различным предметным областям выполняется параллельно, что обеспечивает быструю работу ГА на больших коллекциях документов. Алгоритм редактирования запросов пользователя состоит из трех этапов [12]:

- 1) предварительная обработка текста;
- 2) выделение в документе понятий из онтологии и семантически связанных с ними понятий;
- 3) оценка весовых коэффициентов понятий с помощью ГА.

2. Семантическая индексация документов. Первый этап заключается в проведении морфологического анализа документа: определении начальной формы слов, части речи и т.д. Во время предварительной обработки определяется, встречается ли понятие в заголовке, аннотации или списке ключевых слов.

На втором этапе в тексте определяются те понятия, которые описаны в онтологии предметной области. Остальные слова в документе не учитываются, выполняя роль стоп-слов для алгоритма. Основной сложностью данного этапа является присутствие понятий в выбранной онтологии, которые бы состояли из нескольких слов, т.к. при анализе текста учитываются не отдельные слова, а ряд идущих слов. Для всех понятий в документе формируются характеристики:

1. F – количество повторений понятия в документе.
2. (x_1, x_2, \dots, x_F) – номера позиций, в которых встретилось данное понятие.
3. D – дисперсия понятия, вычисляемая по формуле:

$$D = \frac{\sum_{i=1}^F (x_i - \bar{x})^2}{F},$$

где \bar{x} – среднее значение позиций (x_1, x_2, \dots, x_F) .

Затем необходимо найти в документе семантически близкие понятия к тем, что уже были выделены. Для этого формируется таблица, которая содержит в себе все пары «понятие, вес», где вес – оценка семантической близости (0; 1]. Чем ближе по смыслу понятия, тем выше будет оценка их семантической близости. Семантическую близость можно оценивать различными способами, учитывая количество путей между парами понятий, которые представлены в виде дуг в графе онтологии, а понятия представляются в виде вершин графа, изменение типов дуг в цепочке пути от одного понятия к другому, наличие общих предков и т.д.

Оценка семантической близости понятий для каждого документа происходит только один раз в контексте каждой онтологии предметной области. К моменту завершения формирования онтологического профиля рассматриваемого документа вычисляются все семантические расстояния между парами выделенных понятий. Понятия считаются близкими по смыслу, если их значение семантической близости выше некоторого порогового значения, которые задается в параметрах поиска. Эти понятия также добавляются к списку понятий, которые были найдены в документе. Для понятий, которые не встретились в тексте, частота и дисперсия приравниваются к нулю [13, 14].

3. Методы оценки семантической близости. Существует множество методов оценки семантической близости понятий. Семантическая близость объектов включает в себя множество аспектов сходства, поэтому выбор тех или иных критериев для оценки близости в каждом отдельном случае представляет собой непростую задачу, зависящую от целей исследования. Меры близости онтологических терминов используют разные семантические характеристики сравниваемых понятий – их свойства (различные атрибуты и связи с другими понятиями), взаимное расположение в онтологии и т.д.

Гибридные меры близости являются совокупностью перечисленных мер близости понятий. Чем точнее учитываются характеристики двух понятий с различных точек зрения, тем качественней получается оценка близости этих понятий. В связи с этим именно гибридные меры близости являются наиболее перспективными, т.к. они сочетают в себе сразу несколько подходов оценки близости понятий. В разрабатываемом алгоритме используется гибридная мера оценки близости понятий, основанная на аддитивной свертке:

$$S(c_1, c_2) = \sum_{i=1}^n w_i S^i(c_1, c_2),$$

где S^i – мера близости по выбранному критерию, весовой коэффициент w_i определяет относительную важность критерия, сумма весов равна 1; n – количество используемых критериев. Весовые коэффициенты могут определяются интерактивно экспертами предметной области или пользователями системы, а также автоматически с помощью ГА [15–18].

Сформированный список для начального поиска назовем $ONTO_i$, где i – это количество понятий в списке и $ONTO_i = (c_1, \dots, c_i)$, где c_i – понятие, которое встретилось в документе или близкое по смыслу.

После формирования начального списка понятий необходимо произвести оценку весовых коэффициентов понятий в тексте. Весовой коэффициент понятия – это числовая характеристика, оценивающая выраженность данного понятие в тексте документа. Он зависит от ряда критериев [17]:

- 1) статистических (например, дисперсия и частота);
- 2) онтологических (наличие или отсутствие связей между понятиями);
- 3) структурных (где в документе встречается данное понятие: заголовок, ключевые слова или основной текст).

На старте работы ГА генерирует весовые коэффициенты для всех понятий из начального списка, и во время выполнения генетических операторов (ГО), а также в процессе эволюции, отбираются те понятия, которые удовлетворяют большинству выбранных критериев. В результате генерируется оптимальное распределение весов, которые соответствуют указанным критериям [18].

4. Пространство поиска и генетические операторы. Пространство поиска – множество векторов (w_1, \dots, w_i) длины i , где w_i – это весовой коэффициент понятия c_i из $ONTO_i$ и принимает значение от 0 до 1, i – общее число понятий онтологии, которые семантически связаны с документом.

Хромосомой является вектор (w_1, \dots, w_i) из пространства поиска, а геном является весовой коэффициент одного понятие. Все хромосомы сгенерированы случайным образом, поэтому для оценки их приспособленности используется специальная фитнес-функция и выбраны классические варианты ГО [19]:

1. В скрещивании участвуют p процентов приспособленных особей.
2. Используется одноточечное скрещивание со случайным выбором точки разреза.
3. Во время мутации один ген изменяет свое значение на случайное.

ГА завершает свою работу только после стабилизации популяции, т.е. когда лидирующее значение не меняется в течение нескольких итераций.

Оценка приспособленности хромосом. Значение фитнес-функции указывает насколько точно каждое распределение весовых коэффициентов соответствует рассматриваемому документу. Так как существует большое количество критериев, все они представляются в виде эвристического правила в нотации языка исчисления предикатов первого порядка и будут введены следующие обозначения для описания правил:

- 1) Переменные: a, b – понятия.
- 2) Функции:
 - a) $f(a)$ – количество упоминаний понятия в тексте документа.
 - b) $d(a)$ – дисперсия понятия в документе.
 - c) $c(a, b)$ – семантическая близость между понятиями.
 - d) $w(a)$ – весовой коэффициент понятия.
 - e) $p(a)$ – номер абзаца, где данное понятие встречается впервые.
- 3) Предикаты:

- a) $GT(X, Y)$, $LT(X, Y)$, $GE(X, Y)$, $LE(X, Y)$, $EQ(X, Y)$, $NE(X, Y)$ – стандартные предикаты, обозначающие отношения соответственно $X > Y$, $X < Y$, $X \geq Y$, $X \leq Y$, $X = Y$, $X \neq Y$.
- b) $Close(X, Y, \varepsilon)$ – X близко к Y или $|X - Y| < \varepsilon$.
- c) $Title(a)$, $Annotation(a)$, $Keyword(a)$ – понятие a встречается в различных частях статьи: заголовке, аннотации или списке ключевых слов соответственно.

С учетом установленных интерпретаций предикатных и функциональных символов опишем эвристические правила для оценки весовых коэффициентов понятий:

1) Если понятие a встречается в документе чаще понятие b , следовательно, весовой коэффициент понятия a больше или равен значению весового коэффициента понятия b :

$$\forall a \forall b (GT(f(a), f(b))) \Rightarrow (GE(w(a), w(b))).$$

2) Если значение дисперсии понятия a больше дисперсии понятия b , то весовой коэффициент понятия a больше или равен значению весового коэффициента понятия b :

$$\forall a \forall b (GT(d(a), d(b) \& GT(d(b), 0)) \Rightarrow (GE(w(a), w(b))).$$

3) Если понятие обнаружилось в первом абзаце, то его весовой коэффициент должен быть больше или равен весовым коэффициентам понятий, впервые встречающихся в других абзацах:

$$\forall a \forall b (EQ(p(a), 1) \& \& NE(p(b), 1)) \Rightarrow (GE(w(a), w(b))).$$

4) Весовые коэффициенты синонимов должны быть близки:

$$\forall a \forall b (GT(c(c, b), \beta) \Rightarrow (Close(w(a), w(b), \varepsilon)).$$

5) Если понятия a и b являются синонимами, и сумма их частот равна $f(a) + f(b)$, то весовой коэффициент для этих понятий должен быть больше или равен весовому коэффициенту любого другого понятия c , частота которого равна $f(c) < f(a) + f(b)$:

$$\forall a \forall b \forall c (EQ(c(a, b), 1) \& NE(c(b, c), 1) \& EQ(c(a, c), 1) \& LT(f(c), f(a) + f(b))) \Rightarrow (GE(w(a), w(c)) \& GE(w(b), w(c))).$$

6) Если понятие находится в заголовке, аннотации или в списке ключевых слов, то его весовой коэффициент должен быть больше или равен весовому коэффициенту понятия, которое встречается только в основном тексте документа.

Запросы на естественном языке могут быть сформулированы в произвольном виде: в виде строки поиска по типу поисковых систем *Internet*, в форме фрагмента текстового документа или полного текста, а также в виде списка понятий, которые отмечены в иерархическом отображении понятий из онтологии. Анализ текста запроса производится по аналогии с анализом всего текста искомого документа и представляет запрос в виде:

$$Q = ((cq_1, wq_1), \dots, (cq_n, wq_n)),$$

где cq_i – i -е понятие из запроса пользователя, wq_i – весовой коэффициент понятия cq_i , n – количество выделенных понятий в запросе. Для онтологического профиля каждого документа D_i из полученного результирующего набора документов вычисляется скалярное произведение $\langle Q, D_i \rangle$. Значение данного скалярного произведения является оценкой релевантности документа запросу. Поэтому в итоговый результирующий набор документов попадут только те, значение скалярного произведения для которых выше указанного порогового значения [20].

5. Экспериментальные исследования. Работа алгоритма представлена на примере использования семантической индексации документа в контексте двух небольших предметных онтологий. Первая онтология описывает предметную область для решения задач оптимизации, вторая посвящена искусственным нейронным сетям, рис. 1.

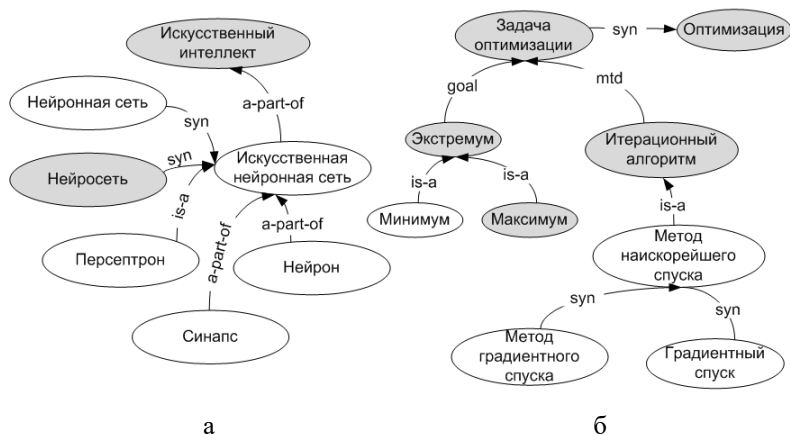


Рис. 1. Фрагменты онтологий: а – искусственных нейронных сетей; б – методов оптимизации

На рисунке цветом выделены понятия, которые не встречаются в тексте статьи. В рассматриваемых онтологиях используются следующие типы связей:

- ◆ *is-a* – отношение класс-подкласс;
- ◆ *a-part-of* – отношение часть-целое;
- ◆ *syn* – отношение синонимии;
- ◆ *goal* – цель;
- ◆ *mtd* – метод.

Сформированные онтологические профили рассматриваемого документа с учетом контекста рассматриваемых онтологий представлены в табл. 1 и 2.

Таблица 1

Онтологический профиль документа в контексте онтологии методов оптимизации

Понятие из онтологии методов оптимизации	Весовой коэффициент
Минимум	0.94
Градиентный спуск	0.62
Итерационный алгоритм	0.45
Максимум	0.38
Оптимизация	0.27
Задача оптимизации	0.15
Экстремум	0.08

Таблица 2

Онтологический профиль документа в контексте онтологии искусственных нейронных сетей.

Понятие из онтологии искусственных нейронных сетей	Весовой коэффициент
Нейрон	0.90
Искусственная нейронная сеть	0.88
Нейронная сеть	0.75
Синапс	0.60
Перцептрон	0.58
Искусственный интеллект	0.55

Во время автоматического построения онтологических профилей документа предполагается, что семантическая близость понятий вычисляется на основе описанной выше гибридной меры оценки близости, при этом расстояние между синонимами равно нулю, весовой коэффициент для типа связей *is-a* равно 0.1, *a-part-of* – 0.3, для всех остальных типов связей – 0.6.

Из приведенного примера можно сделать вывод, что:

1. Онтологический профиль документа отражает наличие взаимосвязанных понятий из онтологии, на базе которой он был сформирован.
2. Наибольший весовой коэффициент имеют понятия, которые встречаются в тексте максимальное количество раз и включаются в заголовок, аннотацию или список ключевых слов.
3. Если понятие не появляется в тексте, но существует семантическая связь с одним из часто встречающихся понятий, то весовой коэффициент данного понятия будет высоким.

Заключение. В работе описан генетический алгоритм, позволяющий проводить индексацию больших коллекций текстовых документов в соответствии с контекстом онтологии. Также представлен алгоритм поиска по построенному семантическому индексу. Использование генетических операторов позволяет получать более точные результаты поиска, а индексация документов сразу по нескольким онтологиям расширяет область поиска. Вычисляемые весовые коэффициенты для оценки меры близости обеспечивают более точную работу семантического поиска за счет индексации не используемых в запросе или документе понятий, но являющихся семантически близкими.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Bova V.V., Kravchenko Y.A., Kureichik V.V. Development of Distributed Information Systems: Ontological Approach // Software Engineering in Intelligent Systems. Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015). Vol. 3. – Springer International Publishing AG Switzerland, 2015. – P. 113-122.
2. Kravchenko Y.A., Kureichik V.V. Knowledge management based on multi-agent simulation in informational systems // Conference proceedings. 8th IEEE International Conference “Application of Information and Communication Technologies – AICT 2014”. – 15-17 October 2014, Astana, Kazakhstan. – P. 264-267.
3. Bova V.V., Kravchenko Y.A., Kureichik V.V. Decision Support Systems for Knowledge Management // Software Engineering in Intelligent Systems. Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015). Vol. 3. – Springer International Publishing AG Switzerland, 2015. – P. 123-130.
4. Amerland D. Google Semantic Search: Search Engine Optimization (SEO) Techniques That Gets Your Company More Traffic, Increases Brand Impact and Amplifies Your Online Presence. – Que Publishing, 2013. – 230 p.
5. Dukkardt, A.N., Lezhebokov, A.A., Zaporozhets, D. Informational system to support the design process of complex equipment based on the mechanism of manipulation and management for three-dimensional objects models // Advances in Intelligent Systems and Computing. – 2015. – Vol. 347. – P. 59-66.
6. Qing He, Xiu-Rong Zhao, Ping Luo, Zhong-Zhi Shi. Combination methodologies of multiagent hyper surface classifiers: design and implementation issues // Second international workshop, AIS-ADM 2007, Proceedings. – Springer Berlin Heidelberg, 2007. – P. 100-113.
7. Кравченко Ю.А., Запорожец Д.Ю., Лежебоков А.А. Способы интеллектуального анализа данных в сложных системах // Известия Кабардино-Балкарского научного центра РАН. – 2012. – № 3 (47). – С. 52-57.
8. Кравченко Ю.А., Бова В.В. Нечеткое моделирование разнородных знаний в интеллектуальных обучающих системах // Открытое образование. – 2013. – № 4 (99). – С. 70-74.

9. Кулиев Э.В., Новиков А.А., Самойлов А.Н., Старкова А.С. Ранжирование онтологий в Semantic Web // Информатизация и связь. – 2016. – № 3. – С. 97-101.
10. Кравченко Ю.А. Синтез разнородных знаний на основе онтологий // Известия ЮФУ. Технические науки. – 2012. – № 11 (136). – С. 141-145.
11. Srikant, R., and Agrawal, R. Mining generalized association rules // Proceedings of VLDB '10. – 2010. – P. 407-419.
12. Тузовский А.Ф., Чуриков С.В., Ямпольский В.З. Системы управления знаниями (методы и технологии) / под общ. ред. В.З. Ямпольского. – Томск: Изд-во НТЛ, 2005. – 260 с.
13. Peat H., and Willet P. The limitations of term co-occurrence data from query expansion in document retrieval systems // Journal of the American Society for Information Science. – 2012. – Vol. 42 (5). – P. 378-383.
14. Davies J., Weeks R., and Krohn U. QuizRDF: Search Technology for the Semantic Web. WWW2002 workshop on RDF & Semantic Web Applications, Proc. WWW2008, Hawaii, USA, 2008.
15. Sheth, A., Bertram, C., Avant, D., Hammond, B., Kochut, K., and Warke, Y. Managing Semantic Content for the Web // IEEE Internet Computing. – 2010. – No. 6 (4). – P. 80-87.
16. Stojanovic N., Struder R., and Stojanovic L. An Approach for the Ranking of Query Results in the Semantic Web // Proc. of ISWC '03 (Sanibel Island, FL, October 2003), SpringerVerlag, 2013. – P. 500-516.
17. Нгуен Б.Н., Тузовский А.Ф. Обзор подходов семантического поиска // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2010. – Т. 2, № 2. – С. 234-237.
18. Bova V.V., Kravchenko Y.A., Kureichik V.V. Decision Support Systems for Knowledge Management // Software Engineering in Intelligent Systems: Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015). Vol. 3. – Springer International Publishing AG Switzerland. – P. 123-130.
19. Кравченко Ю.А., Марков В.В. Онтологический подход формирования информационных ресурсов на основе разнородных источников знаний // Известия ЮФУ. Технические науки. – 2013. – № 7 (144). – С. 116-120.
20. Курейчик В.М. Особенности построения систем поддержки принятия решений // Известия ЮФУ. Технические науки. – 2012. – № 7 (132). – С. 92-98.

REFERENCES

1. Bova V.V., Kravchenko Y.A., Kureichik V.V. Development of Distributed Information Systems: Ontological Approach, *Software Engineering in Intelligent Systems. Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015)*. Vol. 3. Springer International Publishing AG Switzerland, 2015, pp. 113-122.
2. Kravchenko Y.A., Kureichik V.V. Knowledge management based on multi-agent simulation in informational systems, *Conference proceedings. 8th IEEE International Conference "Application of Information and Communication Technologies – AICT 2014"*. 15-17 October 2014, Astana, Kazakhstan, pp. 264-267.
3. Bova V.V., Kravchenko Y.A., Kureichik V.V. Decision Support Systems for Knowledge Management, *Software Engineering in Intelligent Systems. Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015)*. Vol. 3. Springer International Publishing AG Switzerland, 2015, pp. 123-130.
4. Amerland D. Google Semantic Search: Search Engine Optimization (SEO) Techniques That Gets Your Company More Traffic, Increases Brand Impact and Amplifies Your Online Presence. Que Publishing, 2013, 230 p.
5. Dukkardt, A.N., Lezhebokov, A.A., Zaporozhets, D. Informational system to support the design process of complex equipment based on the mechanism of manipulation and management for three-dimensional objects models, *Advances in Intelligent Systems and Computing*, 2015, Vol. 347, pp. 59-66.
6. Qing He, Xiu-Rong Zhao, Ping Luo, Zhong-Zhi Shi. Combination methodologies of multiagent hyper surface classifiers: design and implementation issues, *Second international workshop, AIS-ADM 2007, Proceedings*. Springer Berlin Heidelberg, 2007, pp. 100-113.

7. Kravchenko Yu.A., Zaporozhets D.Yu., Lezhebokov A.A. Sposoby intellektual'nogo analiza dannykh v slozhnykh sistemakh [Methods data mining in complex systems], *Izvestiya Kabardino-Balkarskogo nauchnogo tsentra RAN* [Proceedings of the Kabardino-Balkar scientific centre of the RAS], 2012, No. 3 (47), pp. 52-57.
8. Kravchenko Yu.A., Bova V.V. Nечеткое моделирование разнородных знаний в интеллектуальных обучающих системах [Fuzzy modeling heterogeneous knowledge in intelligent tutoring systems], *Otkrytoe obrazovanie* [Open Education], 2013, No. 4 (99), pp. 70-74.
9. Kuliev E.V., Novikov A.A., Samoylov A.N., Starkova A.S. Ranzhирование онтологий в Semantic Web [The ranking of ontologies in Semantic Web], *Informatizatsiya i svyaz'* [Informatization and Communication], 2016, No. 3, pp. 97-101.
10. Kravchenko Yu.A. Sintez raznorodnykh znaniy na osnove ontologii [The synthesis of diverse knowledge, based on ontologies], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 11 (136), pp. 141-145.
11. Srikant, R., and Agrawal, R. Mining generalized association rules, *Proceedings of VLDB '10*, 2010, pp. 407-419.
12. Tuzovskiy A.F., Chirikov S.V., Yampol'skiy V.Z. Sistemy upravleniya znaniyami (metody i tekhnologii) [The knowledge management system (methods and techniques)], ed. by V.Z. Yampol'skogo. Tomsk: Izd-vo NTL, 2005, 260 p.
13. Peat H., and Willet P. The limitations of term co-occurrence data from query expansion in document retrieval systems, *Journal of the American Society for Information Science*, 2012, Vol. 42 (5), pp. 378-383.
14. Davies J., Weeks R., and Krohn U. QuizRDF: Search Technology for the Semantic Web. WWW2002 workshop on RDF & Semantic Web Applications, Proc. WWW2008, Hawaii, USA, 2008.
15. Sheth, A., Bertram, C., Avant, D., Hammond, B., Kochut, K., and Warke, Y. Managing Semantic Content for the Web, *IEEE Internet Computing*, 2010, No. 6 (4), pp. 80-87.
16. Stojanovic N., Struder R., and Stojanovic L. An Approach for the Ranking of Query Results in the Semantic Web, *Proc. of ISWC '03 (Sanibel Island, FL, October 2003)*, SpringerVerlag, 2013, pp. 500-516.
17. Nguen B.N., Tuzovskiy A.F. Obzor podkhodov semanticheskogo poiska [Overview of the approaches for semantic search], *Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniya i radioelektroniki* [Reports of Tomsk state University of control systems and Radioelectronics], 2010, Vol. 2, No. 2, pp. 234-237.
18. Bova V.V., Kravchenko Y.A., Kureichik V.V. Decision Support Systems for Knowledge Management, *Software Engineering in Intelligent Systems: Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015)*. Vol. 3. Springer International Publishing AG Switzerland, pp. 123-130.
19. Kravchenko Yu.A., Markov V.V. Ontologicheskii podkhod formirovaniya informatsionnykh resursov na osnove raznorodnykh istochnikov znaniy [Ontological approach of formation of information resources based on heterogeneous sources of knowledge], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2013, No. 7 (144), pp. 116-120.
20. Kureychik V.M. Osobennosti postroeniya sistem podderzhki prinyatiya resheniy [Features of construction of systems of support of acceptance of decisions] *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 7 (132), pp. 92-98.

Статью рекомендовал к опубликованию д.т.н., профессор А.Н. Афанасьев.

Марков Владимир Васильевич – Южный федеральный университет; e-mail: vvmarkov@sfedu.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

Новиков Антон Александрович – e-mail: anton.a.novikov@gmail.com; кафедра систем автоматизированного проектирования; аспирант.

Старкова Юлия Сергеевна – e-mail: ysstarkova@sfedu.ru; кафедра систем автоматизированного проектирования; аспирант.

Кравченко Даниил Юрьевич – e-mail: kravchenkodanil122@gmail.com; тел.: 88634371689; кафедра систем автоматического управления; студент.

Markov Vladimir Vasilievich – Southern Federal University; e-mail: vmarkov@sfedu.ru; 44, Nekrasovskiy lane, Taganrog, 347928, Russia; phone: 88634371651; the department of computer aided design; associate professor.

Novikov Anton Alexandrovich – e-mail: anton.a.novikov@gmail.com; the department of computer aided design; postgraduate.

Starkova Yulia Sergeevna – e-mail: ysstarkova@sfedu.ru; the department of computer aided design; postgraduate.

Kravchenko Daniil Yurievich – e-mail: kravchenkodanil122@gmail.com; phone: +78634371689; the department of automatic control systems; student.