

## Автоматическое установление родовидовых отношений между понятиями

**Макарова А.Е.** \*

АО «НПК “ВТ и СС”», г. Москва, Российская Федерация

ORCID: <https://orcid.org/0000-0002-9232-6838>

e-mail: [anna20497@list.ru](mailto:anna20497@list.ru)

**Никитин Ю.В.** \*\*

Институт проблем информатики ФИЦ ИУ РАН

г. Москва, Российская Федерация

ORCID: <https://orcid.org/0000-0002-7641-0247>

e-mail: [yuri.v.nikitin@gmail.com](mailto:yuri.v.nikitin@gmail.com)

**Хорошилов А.А.** \*\*\*

АО «НПК “ВТ и СС”», г. Москва, Российская Федерация

ORCID: <https://orcid.org/0000-0003-4885-3232>

e-mail: [a.a.horoshilov@mail.ru](mailto:a.a.horoshilov@mail.ru)

В работе описываются результаты исследований по созданию методов построения формализованного смыслового описания документов для решения задач семантического поиска документов. Применяемые в исследовании методы базируются на использовании процедур машинной грамматики и концептуального анализа текстов, обеспечивающих выявление его понятийного состава и назначения наименованиям понятий характеристик, соответствующих их семантической роли и значимости в тексте. Для выполнения данной работы был создан комплекс программных средств, который был опробован на документах СМИ.

**Ключевые слова:** семантико-синтаксический анализ, морфологический анализ, концептуальный анализ, автоматическое установление родовидовых отношений.

### Для цитаты:

*Макарова А.Е., Никитин Ю.В., Хорошилов А.А.* Автоматическое установление родовидовых отношений между понятиями // Моделирование и анализ данных. 2022. Том 12. № 1. С. 49–59. DOI: <https://doi.org/10.17759/mda.2022120104>

\**Макарова Анна Евгеньевна*, младший программист, АО «НПК “ВТ и СС”», г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0002-9232-6838>, e-mail: [anna20497@list.ru](mailto:anna20497@list.ru)

\*\**Никитин Юрий Викторович*, научный сотрудник, Институт проблем информатики ФИЦ ИУ РАН, г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0002-7641-0247>, e-mail: [yuri.v.nikitin@gmail.com](mailto:yuri.v.nikitin@gmail.com)

\*\*\**Хорошилов Александр Александрович*, ведущий программист, АО «НПК “ВТ и СС”», г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0003-4885-3232>, e-mail: [a.a.horoshilov@mail.ru](mailto:a.a.horoshilov@mail.ru)



## 1. ВВЕДЕНИЕ

В настоящее время все большую популярность завоевывают технологии семантического полнотекстового поиска, базирующиеся на смысловом анализе информации, в процессе которого производится извлечение смысла из текстовой информации и трансформация его в формальную смысловую модель. Наиболее эффективным лингвистическим инструментом, позволяющим реализовать такую модель поиска, являются так называемые «онтологии» – семантические инструменты, реализующие концептуальные знания о мире в целом и о предметных областях в частности. Описание наиболее общих знаний об окружающем мире содержится в онтологиях верхнего уровня. Знания о конкретных предметных областях содержатся в предметных (тематических) онтологиях.

Следует отметить, что создание онтологических ресурсов требует больших временных затрат и, как показывает практика, их объемы редко превышают несколько десятков тысяч словарных статей. Между тем понятийная система высокотехнологичных отраслей может достигать нескольких сотен тысяч терминологических понятий. Кроме того, в онтологиях наименования понятий представлены в обобщенной форме, а в научно-технических текстах они встречаются в их всевозможных конкретных представлениях, которые не всегда присутствуют в составе конкретной онтологии. Выходом из создавшейся ситуации может быть только автоматизированное создание тематических онтологий большого объема по реальным текстам, относящимся к данной тематике, включая тексты нормативно-технической, проектной и эксплуатационной документации тематических областей.

Построение онтологий – сложный и достаточно длительный процесс, при котором значительные трудовые затраты приходятся на человека-эксперта. Чтобы облегчить его труд, в середине 90-х годов начали создаваться программные среды для разработки онтологий. В состав этих сред были включены интерфейсы, которые позволили выполнять ряд операций построения онтологий: концептуализацию, реализацию, проверку непротиворечивости и документирование. За последние годы число инструментов онтологий резко возросло (сайт консорциума W3C, например, предоставляет список более чем 50 инструментов редактирования).

Обзор методов автоматизированного построения онтологий

Существующие методы автоматического построения онтологий можно условно разделить на несколько групп в зависимости от использования основного подхода:

- методы, основанные на лексико-синтаксических шаблонах;
- методы, основанные на системе продукций;
- статистические методы;
- методы, использующие лингвистические подходы.

*Подход на основе лексико-синтаксических шаблонов.* Данный подход относится к группе методов автоматического построения онтологий, использующих лингвистические средства [1]. Сторонники подхода утверждают, что для построения онтологий следует активно использовать все уровни анализа естественного языка: морфологию, синтаксис и семантику. Подход, основанный на лексико-синтаксических ша-



блогах, давно используется в компьютерной лингвистике. Лексико-синтаксические шаблоны представляют собой характерные выражения и конструкции определенных элементов языка. Данная методика семантического анализа не является ориентированной на определенную предметную область. На основе лексико-синтаксических шаблонов выделяются онтологические конструкции. Ввиду сложности задачи, оценка результатов применения этого подхода проводится опосредованно через анализ результатов его использования, например, в различных приложениях Semantic Web. В целом, лексико-синтаксические шаблоны, как метод семантического анализа данных на естественном языке (в случае большого объема коллекции шаблонов), являются эффективным средством для автоматического построения онтологий.

*Подход на основе системы продукций.* Данный подход относится к группе методов автоматического построения онтологий, в основе которых лежат подходы из области искусственного интеллекта [2]. Предполагается, что эффективное автоматическое построение онтологий может быть основано на способности методов искусственного интеллекта к извлечению из данных элементов знаний и их нетривиальной переработке. Анализ области естественно-языковой обработки информации показывает преобладание использования различных правил при решении задач в рассматриваемой предметной области. Для создания методов автоматического построения онтологий, как правило, разрабатывается модель генерации системы продукций (на основе применения генетического программирования), модель генерации преобразователей (на основе генетического и автоматного программирования), модель генерации систем логического вывода (также на основе генетического и автоматного программирования) и модель аппарата активации продукций (на основе применения автоматного программирования).

*Подход на основе статистических методов.* Этот подход к решению проблемы автоматического построения онтологий базируется преимущественно на статистических методах анализа данных на естественном языке [3]. Для его реализации требуются большие объемы текстовой информации (репрезентативные корпуса текстов). При создании такого корпуса текстов требуется его обязательная предварительная обработка. Процесс такой обработки может быть достаточно трудоемким и обычно состоит из нескольких этапов:

- приведение документов к единому формату;
- токенизация;
- стемминг (лемматизация);
- исключение стоп-слов.
- ручная обработка (с использованием инструментов консорциума W3C) по установлению смысловых связей между терминологическими наименованиями понятий.

Однако не всегда есть необходимость в проведении всех вышеперечисленных этапов. В результате предварительной обработки каждый документ коллекции характеризуется вектором типов данного документа и их частотой встречаемости. На первом этапе построения онтологии выделяются входящие в ее состав классы, которые, как правило, базируются на терминах предметной области.

Таким образом, основная задача построения онтологий сводится к выявлению терминов рассматриваемой предметной области. Алгоритмы извлечения терминов из



текстов на естественном языке можно разделить на две группы: статистические и лингвистические. Однако первые обладают определенным преимуществом, поскольку их использование не зависит от лингвистических особенностей конкретного языка. Существующие статистические методы могут показать лучшие результаты, если дополнить их определенными эвристиками.

В качестве базовых эвристик предлагается использовать следующие:

*Эвристика № 1.*

Имя класса содержит хотя бы одно существительное.

*Эвристика № 2.*

Общепотребительные слова по сравнению с терминами обладают большей частотой встречаемости, приблизительно равной в различных предметных областях.

*Эвристика № 3.*

Считается, что количество информации термина из нескольких слов больше, чем количество информации отдельных слов, входящих в его состав.

*Статистический подход* базируется на определенной технологической схеме. На первом этапе в каждой коллекции документов выделяют существительные, и определяют их частоту встречаемости. При этом использование частотных критериев значительно сокращает число предполагаемых классов понятий. На втором этапе выделяют термины, состоящие из одного слова. На основании выдвинутой эвристики № 1 сравниваются частоты встречаемости различных существительных в рамках одной коллекции, а также проводится оценка пересечения различных коллекций по используемым существительным (эвристика № 2). Однако статистические данные – не единственный источник классов онтологии.

Терминологические словари также могут стать источниками знаний при формировании ядра онтологии. В случае работы с коллекциями неспециализированных в конкретной области документов возможно использование существующих разработанных экспертами онтологий (например, для английского языка – онтологии WordNet). Наконец, на третьем этапе на основе взаимной информации могут быть выделены термины, состоящие из нескольких слов. Стоит отметить, что в данном случае используется эвристика № 3. Для случая двухсложных терминов получаем, что взаимная информация определяется по формуле:

$$mi(x,y) = P(x,y) / P(x)P(y), \quad (1)$$

где  $x$  и  $y$  представляют собой отдельные слова термина,

$P(x)$  – частота встречаемости  $x$ ,

$P(y)$  – частота встречаемости  $y$ ,

$P(x,y)$  – частота совместной встречаемости  $x$  и  $y$ .

Выделенные описанным выше образом термины будут представлять собой классы будущей онтологии [2, 3].

Автоматическое формирование тематических словарей на основе синтагматических шаблонов синтагм.

Существенным недостатком всех выше рассмотренных методов создания онтологий требуют значительных трудозатрат экспертов высокой квалификации на раз-



личных этапах этого процесса, что не позволяет построить в приемлемые сроки онтологические ресурсы большого объема. Решить эту проблему возможно только с привлечением достаточно мощных средств автоматизации этого процесса. К таким средствам можно отнести набор базовых средств смыслового анализа текстов: морфологический, семантико-синтаксический и концептуальный анализ текстов.

Авторы настоящего исследования располагали такими средствами [4, 5].

В качестве основного инструмента авторами использовался морфологический анализатор МетаФраз [6], семантико-синтаксический анализатор МетаФраз [7] и разработанная авторами настоящей статьи процедура концептуального анализа текстов, базирующаяся на синтагматических шаблонах. Положенные в основу метода этой процедуры синтагматические шаблоны функционируют на основе метода лингвистической аналогии текстовых форм представлений наименований понятий в текстах и по сути представляют собой синтаксическую модель фрагмента текста в рамках следующего утверждения: представление синтаксической структуры фрагментов текстов в виде последовательности контактно расположенных двухбайтовых индексов элементов синтагм, обладающих грамматическими свойствами конкретных слов-эталонов, позволяет фиксировать грамматические и синтаксические свойства различных последовательностей реальных текстов, а также позволяет в ряде задач распознавать аналогичные по заданным свойствам последовательности слов и словосочетаний в текстах [7].

В качестве эталонных словосочетаний были использованы элементы эталонного концептуального словаря наименований понятий (ЭКСП) объемом 1.7 млн. словарных статей. В таблице 1 приведен фрагмент понятийной системы текста по медицинской тематике, выявленный на основе обобщенных синтагм, полученных по словарю ЭКСП.

Таблица 1

**Фрагмент понятийной системы текста,  
выявленный на основе обобщенных синтагм**

HA Aw HA Aw AB Nw	острого респираторного синдрома / атомного ракетного крейсера
HA Aw AB Nw	респираторного синдрома / абсолютного принципа
HI Aw w m Nw	атипичная пневмония / абдоминальная аортография
HT Aw AANw	воспалительный процесс / абразивный износ
HT Aw AANw HA Aw AB Nw	воспалительный процесс пищеварительного тракта / преступный акт пещерного садизма
AANw HA Aw AB Nw	процесс пищеварительного тракта / анализ клеточного цикла
HA Aw AB Nw	пищеварительного тракта / абсолютного принципа
HB Aw w f Nw	коронавирусной инфекции / аварийной ситуации
HI Aw HI Aw	острая респираторная / вегетативная нервная

Словарная запись в таблице 1 каждого элемента состоит из трех компонент: шаблона синтаксической синтагмы словосочетания (первая слева запись), далее в центре выделенное из текста словосочетание, и последнее справа словосочетание (через косую черту) – словосочетание эталонного словаря.

Автоматизированное составление словарей терминологических наименований понятий по текстам документов можно выполнить по следующей технологической схеме:



1. Формально-логический контроль текстов;
2. Членение исходного текста на отдельные слова;
3. Морфологический анализ слов текста;
4. Членение текста на предложения;
5. Семантико-синтаксический анализ текстов;
6. Приближенный концептуальный анализ текстов;
8. Автоматическое приведение наименований понятий к их канонической форме;
9. Формирование частотного словаря наименований понятий.

## 2. АЛГОРИТМ УСТАНОВЛЕНИЯ РОДОВИДОВЫХ ОТНОШЕНИЙ В ТЕКСТАХ

Полученный понятийный словарь слов и словосочетаний текста достаточно большого объема или совокупной тематической коллекции текстов дает возможность автоматически устанавливать смысловые отношения между словосочетаниями на основе анализа лексического состава словосочетаний. Для установления родовидовых отношений между словосочетаниями можно воспользоваться следующей гипотезой: если имеются два словосочетания различной длины, в которых имеются одинаковые главные (опорные) слова, и все определители более короткого словосочетания совпадают с определителями более длинного словосочетания, то эти два словосочетания находятся в отношении родовидовой связи, и более короткое словосочетание является родовым понятием, а более длинное – видовым;

В соответствие с этой гипотезой был разработан следующий алгоритм установления родовидовых отношений в текстах между наименованиями понятий:

Устанавливаем последовательность слов и словосочетаний, имеющих одно и то же главное слово. Обрабатываемое словосочетание располагается справа от разделителей звездочек, а его главное слово слева. Результаты занесем в таблицу 2.

Таблица 2

### Список словосочетаний, использованных для исследования, и их главное слово

академия***Академии наук
академия***Академия Генштаба
академия***Киево-Могилянская академия
академия***Российская Академия
академия***академия МВД
академия***военная академия
академия***сельскохозяйственная академия
академия***Академия Народного Хозяйства
академия***Академия военных наук
академия***Академия наук Татарстана
академия***Военная академия Генштаба
академия***Польской академии наук
академия***Российская академия народного хозяйства
академия***Украинская свободная академия наук



Строим матрицу  $M$  вхождения отдельных слов в словосочетания. Номера строк матрицы  $i$  – это индексы предложений (или словосочетаний)  $i=1, 2, \dots, n$ , где  $n$  – количество обрабатываемых предложений или словосочетаний. Номера столбцов матрицы  $j$  – индексы уникальных слов в тексте  $j=1, 2, \dots, k$ , где  $k$  – количество всех уникальных слов.

Для заполнения используются нули и единицы. Если в предложении  $i$ , присутствует слово с индексом  $j$ , в ячейке  $a_{ij}$  ставится единица, в противном случае ноль.

$$M = \begin{pmatrix} 1 & \dots & a_k \\ \vdots & \ddots & \vdots \\ a_n & \dots & a_{nk} \end{pmatrix}$$

1. Последовательно устанавливаем вхождения  $N$  слов в словосочетания, начиная с  $N = 1$ . Первым в обработку поступает слово, встречающееся чаще всего, то есть главное. В нашем примере это «академия».
2. Если все слова словосочетания имеют частоту равную единице, то это будут словосочетания 1-го уровня RV отношений. Родовым понятием для них будет одно это слово (если такого родового слова нет, то его необходимо создать принудительно). Приведем пример словосочетаний, относящихся к первому уровню RV отношений, выбрав их из Таблицы 2:

- Академии наук
- Академия Генштаба
- Киево-Могилянская академия
- Российская Академия
- академия МВД
- военная академия
- сельскохозяйственная академия
- Академия Народного Хозяйства

3. Далее в соответствии с матрицей отдельных слов в словосочетаниях устанавливаются отношения следующего уровня RV отношений ( $N+1$ ), путем выявления вхождения слов в словосочетания, имеющих частоту  $f > 1$ .

Примером словосочетаний с RV отношениями второго уровня можно считать:

Для родового понятия «академия наук»:

- Академия военных наук
- Академия наук Татарстана
- Польской академии наук
- Украинская свободная академия наук

Для родового понятия «академия Генштаба»:

- Военная академия Генштаба

Для родового понятия «военная академия»:

- Академия военных наук
- Военная академия Генштаба

Для родового понятия «Российская Академия»:

- Российская академия народного хозяйства



4. Для каждой цепочки устанавливаем словосочетание с минимальной длиной и полностью входящее в другое словосочетание. Для каждой пары продолжаем устанавливать следующий уровень RV отношений.

Заметим, что в нашем примере возможно выделить квазиродовое понятие «академия народного хозяйства», входящее в состав словосочетания «Российская академия народного хозяйства». Они связаны RV отношениями третьего уровня.

5. Процесс установления RV отношений заканчивается, когда будут исчерпаны все понятия с частотной лексикой.

В таблице 3 приведены результаты работы алгоритма.

Таблица 3

### Результаты работы алгоритма

академия =RV= Академии наук
академия =RV= академия МВД
академия =RV= Киево-Могилянская академия
академия =RV= сельскохозяйственная академия
академия =RV= Академия Народного Хозяйства
академия =RV= Академия Генштаба
академии наук =RV= Академия военных наук
академии наук =RV= Польской академии наук
академии наук =RV= Украинская свободная академия наук
академии наук =RV= Академия наук Татарстана
академия Генштаба =RV= Военная академия Генштаба
военная академия =RV= Академия военных наук
военная академия =RV= Военная академия Генштаба
российская Академия =RV= Российская академия народного хозяйства
академия Народного Хозяйства =RV= Российская академия народного хозяйства

## 3. ЗАКЛЮЧЕНИЕ

В проведенном исследовании показана принципиальная возможность на основе предлагаемых методов автоматически создать программные и декларативные средства для процедур автоматического семантического поиска документов, в частности, для процедур автоматического установления степени смысловой близости документов. На основе созданного в процессе проведения настоящего исследования алгоритма автоматического установления родовидовых отношений между понятиями был проведен эксперимент по обработке массива текстов СМИ большого объема. В ходе эксперимента получены удовлетворительные результаты. Их анализ показал, что на количественные характеристики результатов обработки текстов незначительно повлияли ресурсные ограничения и принятые в связи с этим принятые допущения.

Для улучшения работы используемых процедур необходимо предпринять следующие шаги:

1. Необходимо выполнить в полном объеме комплекс технологических операций по созданию декларативных средств. Исходные тексты предварительно должны быть



подвергнуты обработке процедурами формально-логического контроля и исправления орфографических и синтаксических ошибок. Большая часть процедур анализа и исправления текстовых искажений должны быть автоматизированы.

2. Необходимо обеспечить требуемые параметры концептуальных словарей (по объему и составу). При этом необходимо исходить из следующих рекомендаций: тематический словарь должен быть составлен по актуальным текстам и иметь объем не менее 1 млн. словарных статей. Покрытие анализируемых текстов наименованиями понятий должно быть не менее 60–70 % от их общего состава
3. Для повышения степени обобщения смыслового содержания наименований понятий необходимо, чтобы наименования понятий в словарях были связаны также отношениями синонимии.

### *Литература*

1. *Захарова И.В.* Математическая модель семантического поиска с использованием онтологического подхода: Автореф. дис. канд. физ.-мат. наук. – Челябинск, 2009. – 20 с.
2. *Найханова Л.В.* Методы и модели автоматического построения онтологий на основе генетического и автоматного программирования: Автореф. дис. докт. тех. наук. – Красноярск, 2008. – 36 с.
3. *Рабчевский Е.А.* Автоматическое построение онтологий. // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. – СПб.: Издательство Политехнического Университета, 2007. – № 52–2. – С. 22–26.
4. *Хорошилов А.А.* Методы автоматического установления смысловой близости документов на основе их концептуального анализа // Труды XV Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2013, г. Ярославль, 14–17 октября 2013 года. – 2013. – С. 369–376.
5. *Хорошилов А.А., Макарова А.Е.* Автоматизированное формирование онтологических ресурсов в авиакосмической отрасли // Тезисы. – М.: Издательство «Перо», 2021– 9,43 Мб. [Электронное издание]. – Москва: Московский авиационный институт (национальный исследовательский университет), 2021. – С.282–283.
6. Морфологический анализатор МетаФраз нового поколения / Хорошилов Ал-др А., Никитин Ю.В., Пшеничный С.И., Шевкунов М.А., Хорошилов Ал-ей А. // Научно-техническая информация. Серия 2: Информационные процессы и системы. ВИНТИ РАН. 2021. № 5.
7. Автоматическое формирование синтаксической модели языка для задач машинного перевода и информационного поиска / Кан А.В., Ревина В.Д., Руснак В.И., Хорошилов Ал-др А., Хорошилов Ал-сей А. // Сб. «Научно-техническая информация», Сер. 2, № 12, ВИНТИ, 2018.



## Automatic Establishment of Generic Relationships Between Concepts

**Anna E. Makarova\***

AO "NPK "VT i SS"", Moscow, Russia

ORCID: <https://orcid.org/0000-0002-9232-6838>

e-mail: [anna20497@list.ru](mailto:anna20497@list.ru)

**Yuriy V. Nikitin\*\***

Institute for Informatics Problems FITS IU RAN, Moscow, Russia

ORCID: <https://orcid.org/0000-0002-7641-0247>

e-mail: [yuri.v.nikitin@gmail.com](mailto:yuri.v.nikitin@gmail.com)

**Alexsander A. Khoroshilov\*\*\***

AO "NPK "VT i SS"", Moscow, Russia

ORCID: <https://orcid.org/0000-0003-4885-3232>

e-mail: [a.a.horoshilov@mail.ru](mailto:a.a.horoshilov@mail.ru)

The article describes the results of research on the creation of methods for constructing a formalized semantic description of documents for solving problems of semantic search for documents. The methods used in the study are based on the use of machine grammar procedures and conceptual analysis of texts, which ensure the identification of its conceptual composition and the assignment of characteristics to the names of concepts that correspond to their semantic role and significance in the text. To perform this work, a set of software tools was created, which was tested on media documents.

**Keywords:** semantic-syntactic analysis, morphological analysis, conceptual analysis, automatic establishment of generic relations.

### For citation:

Makarova A.E., Nikitin Yu.V., Khoroshilov A.A. Automatic Establishment of Generic Relationships Between Concepts. *Modelirovanie i analiz dannykh = Modelling and Data Analysis*, 2022. Vol. 12, no. 1, pp. 49–59. DOI: <https://doi.org/10.17759/mda.2022120104> (In Russ., abstr. in Engl.).

### References

1. Zakharova I.V. Matematicheskaya model' semanticheskogo poiska s ispol'zovaniem ontologicheskogo podkhoda: Avtoref. dis. kand. fiz.-mat. nauk. – Chelyabinsk, 2009. – 20 p.
2. Naikhanova L.V. Metody i modeli avtomaticheskogo postroeniya ontologii na osnove geneticheskogo i avtomatnogo programirovaniya: Avtoref. dis. dokt. tekhn. nauk. – Krasnoyarsk, 2008. – 36 p.

\***Anna E. Makarova**, Junior programmer, AO "NPK "VT i SS"", Moscow, Russia, ORCID: <https://orcid.org/0000-0002-9232-6838>, e-mail: [anna20497@list.ru](mailto:anna20497@list.ru)

\*\***Yuriy V. Nikitin**, Researcher, Institute for Informatics Problems FITS IU RAN, Moscow, Russia, ORCID: <https://orcid.org/0000-0002-7641-0247>, e-mail: [yuri.v.nikitin@gmail.com](mailto:yuri.v.nikitin@gmail.com)

\*\*\***Alexsander A. Khoroshilov**, Senior programmer, AO "NPK "VT i SS"", Moscow, Russia, ORCID: <https://orcid.org/0000-0003-4885-3232>, e-mail: [a.a.horoshilov@mail.ru](mailto:a.a.horoshilov@mail.ru)



3. Rabchevskii E.A. Avtomaticheskoe postroenie ontologii. Nauchno-tekhnicheskie vedomosti Sankt-Peterburgskogo gosudarstvennogo politekhnicheskogo universiteta. – SPb.: Izdatel'stvo Politekhnicheskogo Universiteta, 2007. – № 52–2. – P. 22–26.
4. Khoroshilov A.A. Metody avtomaticheskogo ustanovleniya smyslovoi blizosti dokumentov na osnove ikh kontseptual'nogo analiza. Trudy XV Vserossiiskoi nauchnoi konferentsii «Elektronnye biblioteki: perspektivnye metody i tekhnologii, ehlektronnye kolleksiil» – RCDL'2013, g. Yaroslavl', 14–17 oktyabrya 2013 goda. – 2013. – P. 369–376.
5. Khoroshilov A.A., Makarova A.E. Avtomatizirovannoe formirovanie ontologicheskikh resursov v aviakosmicheskoi otrasli. Tezisy. – M.: Izdatel'stvo «PerO», 2021. – 9,43 Mb. [Elektronnoe izdanie]. – Moskva: Moskovskii aviatsionnyi institut (natsional'nyi issledovatel'skii universitet), 2021. – P. 282–283.
6. Morfologicheskii analizator MetAFraz novogo pokoleniya / Khoroshilov Al-dr A., Nikitin YU.V., Pshenichni S.I., Shevkunov M.A., Khoroshilov Al-ei A. Nauchno-tekhnicheskaya informatsiya. Seriya 2: Informatsionnye protsessy i sistemy. VINITI RAN. 2021. № 5.
7. Avtomaticheskoe formirovanie sintaksicheskoi modeli yazyka dlya zadach mashinnogo perevoda i informatsionnogo poiska. Kan A.V., Revina V.D., Rusnak V.I., Khoroshilov Al-dr A., Khoroshilov Al-sei A. Sb. «Nauchno-tekhnicheskaya informatsiya», Ser. 2, № 12, VINITI, 2018.

Получена 18.02.2022

Принята в печать 14.03.2022

Received 18.02.2022

Accepted 14.03.2022