УДК 004.891
DOI: 10.15827/0236-235X.139.420-427
Дата подачи статьи: 08.06.22
2022. Т. 35. № 3. С. 420–427

Автоматизированная система анализа ключевых терминов

С.А. Власова [1], *к.т.н., ведущий научный сотрудник, svlasova@jscc.ru*
Н.Е. Каленов [1], *д.т.н., профессор, главный научный сотрудник, nkalenov@jscc.ru*

[1] *Межведомственный суперкомпьютерный центр (МСЦ) РАН – филиал ФНЦ Научно-исследовательский институт системных исследований (НИИСИ) РАН, г. Москва, 119991, Россия*

В статье описан предложенный авторами метод формирования массива ключевых терминов, составляющих основу предметной онтологии Единого цифрового пространства научных знаний и перечня статей (слотов) энциклопедий по определенному научному направлению.

Метод основан на частотном анализе встречаемости ключевых терминов в статьях, опубликованных в ведущих научных журналах по данной тематике. Методика предполагает программную обработку метаданных статей, отраженных в различных БД, построение рейтинговых списков частоты встречаемости отдельных ключевых терминов и выделение ядер таких списков, что, в свою очередь, можно рассматривать как основу для наполнения предметной онтологии и формирования энциклопедических слотов.

Для реализации методики авторами данной статьи разработаны структура соответствующей БД, программные средства для ее наполнения, обработки и анализа данных. В статье представлены БД и результаты практической реализации методики на основе обработки нескольких тысяч статей из ведущих российских журналов по математике, информатике и физике (выявлены и проанализированы термины на русском и английском языках).

Проведенная оценка соответствия частот распределения ключевых терминов и составляющих их отдельных слов закону Брэдфорда показала значительные расхождения с этим законом в случае ключевых терминов, но определенные сближения при рассмотрении отдельных слов и их пермутации внутри ключевых терминов.

Ключевые слова: *ключевой термин, метаданные статей, частотный анализ, БД, автоматизированная система.*

В число наиболее важных направлений, связанных с сохранением и распространением достижений науки в России, входят развитие *Большой российской энциклопедии* (БРР) (https://bigenc.ru), создание на ее основе портала «Знание» и формирование *единого цифрового пространства научных знаний* (ЕЦПНЗ) как интегратора разнородных научных информационных ресурсов [1, 2].

Основу любой энциклопедии составляют статьи (слоты), посвященные персонам, событиям, научным направлениям и т.п. Определение перечня слотов является одной из важных проблем при формировании энциклопедий. Подобная же проблема, но в несколько другом ракурсе, возникает при создании ЕЦПНЗ, терминологической основой которого являются предметные онтологии, описывающие конкретное научное направление в виде структурированных терминов и их связей. Совокупность предметных онтологий и поддерживающая их программная оболочка ЕЦПНЗ должны обеспечивать развитый многоаспектный поиск разнородной информации, ее удобную визуализацию и навигацию по связанным ресурсам. Проблема формирования предметных онтологий тесно связана с традиционными задачами формирования тематических тезаурусов [3, 4]. Основой тезауруса, предметной онтологии, а также «сырьем» для определения перечня энциклопедических слотов в научной области должны стать словники, включающие перечень терминов, описывающих то или иное научное направление.

Соответственно, возникает задача формирования таких словников, включающих, с одной стороны, наиболее важные широко распространенные термины, а с другой – термины, описывающие новые перспективные направления, определяющие развитие конкретной научной области. Эту задачу было предложено решать с помощью *ключевых терминов* (КТ), используемых авторами в своих научных статьях (далее – авторские КТ) [5, 6].

Анализ современных публикаций, связанных с понятием «ключевой термин», показал, что ни в одной из них не исследуется задача анализа массива КТ по определенному научному направлению. Авторы предлагают различные методы извлечения КТ из текстовых документов [7–9], исследуют этимологию и историю становления ключевых понятий в определенной области [10, 11], анализируют возникновение и введение в научный оборот новых терминов [11, 12], проводят исследования по использованию и сравнению определенных терминов разных языков [13, 14].

Многоаспектный анализ авторских КТ предполагает возможность выявления общего ядра – перечня наиболее часто употребляемых КТ, анализ временной динамики частоты появления тех или иных КТ, возможность связывать их с определенными журналами и статьями и т.п. Для решения этих задач необходимо иметь БД, содержащую соответствующие элементы, и программный инструментарий, позволяющий обрабатывать их и визуализировать результаты обработки.

Структура БД

Поддерживаемая Microsoft SQL Server БД системы содержит семь видов объектов: тематика, журнал, статья, КТ на русском и английском языках, *ключевые слова* (КС) (отдельные слова, входящие в состав терминов) на русском и английском языках. Объект каждого вида имеет следующие атрибуты.

Тематика: идентификатор записи, наименование тематики журнала, рубрика ГРНТИ журнала.

Журнал: идентификатор записи, название журнала на русском (английском) языке.

Статья: идентификатор записи, название статьи на русском (английском) языке, идентификатор журнала, год издания, выпуск (том, номер), адрес сайта статьи.

КТ на русском (английском) языке: идентификатор записи, КТ на русском (английском) языке, идентификатор статьи, идентификатор журнала.

КС на русском (английском) языке: идентификатор записи, КС на русском (английском) языке, идентификатор КТ.

Программная оболочка, обеспечивающая работу с БД, создана на основе технологии Microsoft ASP.NET на платформе Microsoft .NET Framework в среде разработки Microsoft Visual Studio 2019.

Отбор материалов для анализа

Для получения устойчивых результатов, отражающих реальное распределение КТ, необходимо иметь репрезентативную выборку статей по рассматриваемому научному направлению и, соответственно, достаточно большой массив журналов, содержащих в цифровом виде информацию о КТ.

Для проведения соответствующих расчетов, касающихся русскоязычных терминов, наиболее рационально было бы использовать БД РИНЦ или RSCI. Однако РИНЦ не дает возможности выгрузки статей в структурированном виде и анализа HTML-файлов, выдаваемых по запросам статей. За предоставление возможности анализа массива данных РИНЦ самим пользователям администрация eLibrary требует постатейную достаточно высокую плату. БД RSCI, представленная на платформе Web of Science и содержащая, как утверждается, наиболее важные российские журналы, в национальной подписке для российских пользователей недоступна, для работы с ней необходимо коммерческое соглашение с компанией Clarivate.

Поэтому для проведения модельных расчетов авторами данного исследования была выбрана отечественная система MathNet, которая позволяет анализировать поддерживаемую ею информацию программно. В дополнение к этому были проанализированы сайты журналов, не отражаемых в MathNet, на предмет возможности программного выделения КТ из метаданных опубликованных в них статей.

В результате были отобраны перечисленные далее журналы по математике, физике и информатике.

Математика: «Известия Российской академии наук. Серия математическая», «Математический сборник», «Дискретная математика», «Успехи математических наук», «Функциональный анализ и его приложения», «Алгебра и анализ», «Алгебра и логика», «Труды Математического института им. В.А. Стеклова РАН».

Физика: Вестник Самарского государственного технического университета. Серия «Физико-математические науки», «Теоретическая и математическая физика».

Информатика: «Вычислительные методы и программирование», «Программные продукты и системы», «Информатика и ее применения».

По математике было выбрано для анализа более 6 000 статей, по физике – около 3 500, по

информатике – более 3 500. Временной охват статей по каждому направлению составил в среднем 16 лет.

Загрузка информации в БД системы

На портале Math-Net.Ru представлен список названий журналов (http://www.mathnet.ru/ej.phtml?option_lang=rus), из которого и был сделан выбор для загрузки в систему. Названия журналов являются активными ссылками, переход по которым приводит к странице выбранного журнала. Войдя в раздел «Архив», получаем список всех выпусков (номеров) журнала с указанием тома и года. Для получения статей переходим по ссылке соответствующего номера. Названия статей являются активными ссылками, которые приводят к странице с ее данными. Например, на рисунке 1 отображены все данные, необходимые для загрузки в систему: название журнала, название статьи, год выпуска, КС. Для получения этих данных на английском языке нужно в адресе статьи изменить значение параметра option_lang на eng. Параметр papered определяет идентификатор конкретной статьи. Для загрузки данных статей в БД системы была разработана специальная программа, выполняющая следующие действия.

Название журнала на русском и английском языках загружается в БД системы. Последовательно, изменяя значение параметра papered, определяется наличие страницы со статьей на портале Math-Net.Ru (на русском языке option_lang=rus, на английском – option_lang=eng). Для каждой найденной статьи проводится анализ ее HTML-кода: определяются год, название статьи, КС.

Год публикации отделен запятой от названия журнала после тега <a title='. Например,

```
<a title='Дискретая математика, 2017,
том 29,
```

Название журнала заключено в теги </size>. Например,

```
<span class=red><font size=+1>
Линейно реализуемые автоматы</font>
</span>
```

КС находятся после тега Ключевые слова: в тегах . Например,

```
<b>Ключевые слова:</b>
<i>теория автоматов, переходные систе-
мы, подстановка, кодирование, слож-
ность.</i>
```

В HTML-коде страницы с данными статьи на английском языке ее название заключено в те же теги, что и на русском. Например,

```
<span class=red><font size=+1>Linearly
realizable automata</font></span>
```

КС находятся после тега Keywords: в тегах . Например,

```
<b>Keywords:</b>
<i>automata theory, automata, semiau-
tomata, transition systems, permuta-
tion, substitution function, assign-
ment, state encoding, complexity.</i>
```

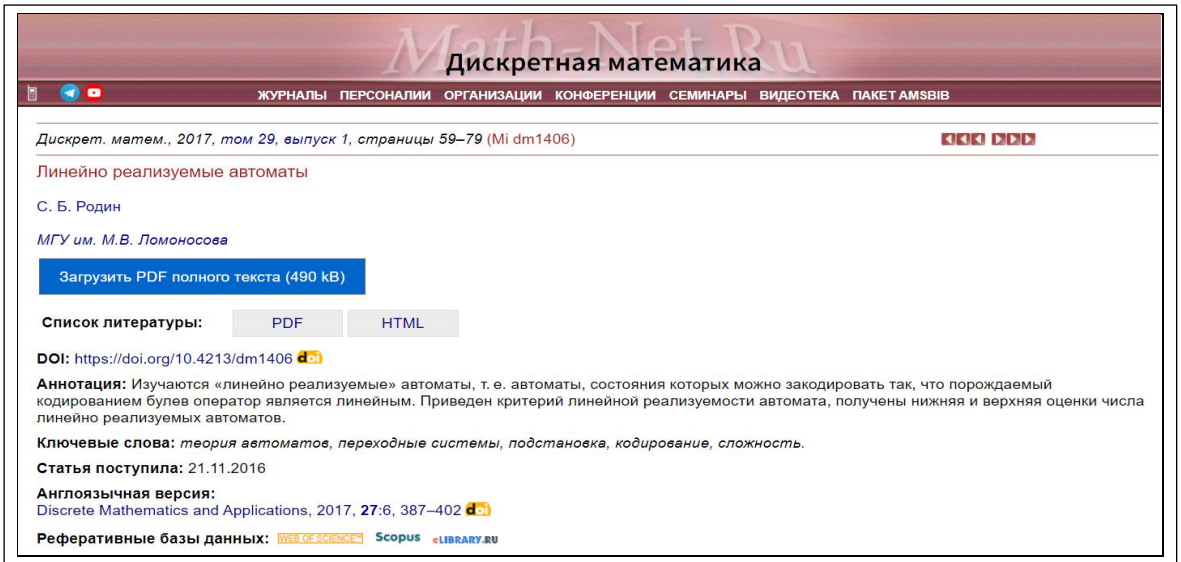

Рис. 1. Пример статьи на портале Math-Net.Ru

Fig. 1. An example of an article on the portal Math-Net.Ru

Разработанная программа реализует алгоритм выбора необходимых метаданных и загружает их в БД описанной выше структуры.

После окончания загрузки данных всех статей выбранного журнала программа из загруженных записей объектов «Ключевой термин на русском языке» и «Ключевой термин на английском языке» формирует данные для объектов «Ключевое слово на русском языке» и «Ключевое слово на английском языке» соответственно.

Алгоритм анализа данных

Для формирования словника, принимаемого в качестве основы предметной онтологии, из массива КТ, содержащего тысячи элементов, необходимо определить критерии отнесения к нему того или иного термина. В качестве основы для такого критерия можно принять частоту встречаемости КТ в выделенном массиве (в данном случае в журналах определенной тематики). Для реализации этого критерия на первом этапе необходимо построить и проанализировать рейтинговый список КТ, упорядоченный по частоте их встречаемости. На следующем этапе нужно определить, по какому принципу из рейтингового списка отбирать КТ для включения в словник. Одним из возможных критериев включения КТ в словник может служить его принадлежность к ядру рейтингового списка, включающего КТ, на которые приходится 80 % общей встречаемости. В соответствии с принципом Парето (который называют еще законом 20/80) [15], ядро должно включать 20 % КТ, находящихся в верхней части рейтингового списка. К числу функций описываемой системы относятся построение ядер массивов КТ и КС и оценка уровня соответствия их распределения принципу Парето.

Интерфейс системы

Система размещена в свободном доступе по адресу http://dirsmsc.ru/keyterms/ и предоставляет пользователю следующие возможности:

– анализ общего частотного распределения КТ и КС;

– хронологический анализ распределения КТ и КС по журналам – по выбранным из ядра КТ или КС можно получить их частотное распределение по годам, а также список журналов, в которых они встречаются (с указанием количества по годам);

– анализ КТ и КС, относящихся к конкретным журналам, – по выбранным журналам можно получить списки ядра КТ и КС (с указанием их встречаемости).

Рассмотрим подробнее возможность проведения анализа в каждом разделе.

Анализ общего частотного распределения КТ и КС. Для проведения данного анализа система предлагает заполнить форму запроса. Из предложенных списков необходимо выбрать тематику и атрибуты, по которым проводится анализ; можно указать временной интервал (по умолчанию все годы), порядок сортировки результирующих списков (по встречаемости или по алфавиту) и выбрать количество КТ (КС) из ядра, необходимое для показа пользователю (по умолчанию 50).

По составленному запросу система выдаст следующие данные: количество всех найденных элементов (КТ или КС), соответствующих запросу, количество различных терминов, количество терминов в ядре, количество повторяющихся терминов ядра, частота встречаемости терминов ядра (процент повторяющихся терминов ядра). Также система выдаст список элементов с указанием частоты встречаемости.

На рисунке 2 приведен пример результата выполнения запроса для анализа КТ в статьях журналов, относящихся к тематике «Математика», опубликованных в период 2000–2021 гг.

Хронологический анализ распределения КТ и КС по журналам. Для проведения данного анализа в предложенной системой форме необходимо выбрать тематику, указать атрибут «Ключевые термины» или «Ключевые слова», выбрать атрибут «Русские» или «Английские», указать количество КТ (КС) из ядра, необходимое для показа пользователю. В соответствии с запросом система выдаст на экран пользователя список КТ (КС), отсортированных в порядке частоты встречаемости. На рисунке 3 приведен пример запроса на поиск русских КТ по тематике «Информатика».

Пользователь из предоставленного списка выбирает нужные КТ. Для каждого выбранного КТ система выдаст его частотное распределение по годам, а также список журналов, в которых он встречается. Для каждого журнала также будет выдано частотное распределение термина по годам (см. http://www.swsys.ru/uploaded/image/2022-3/2022-3-dop/7.jpg).

Анализ КТ и КС в выбранных журналах. В данном разделе система предлагает выбрать тематику, после чего выдает список журналов

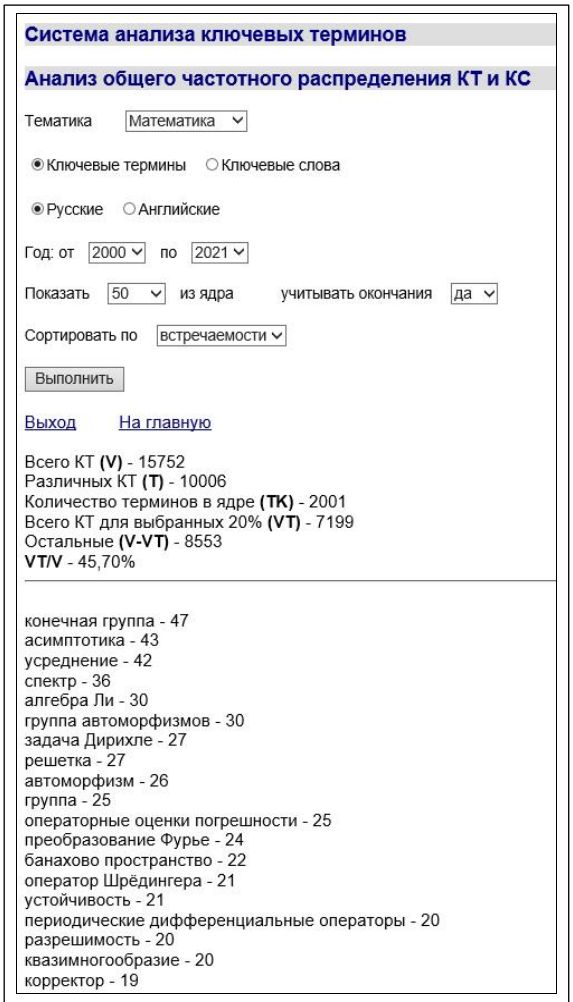

Рис. 2. Пример анализа частотного распределения КТ

Fig. 2. An example of analyzing the frequency distribution of key terms

по указанной тематике. В этом списке пользователю нужно отметить интересующие его журналы и задать необходимые данные для выбора и показа результатов (см. http://www.swsys.ru/uploaded/image/2022-3/2022-3-dop/8.jpg). Система, обработав запрос, выдаст список найденных КТ с указанием их встречаемости в журнале (см. http://www.swsys.ru/uploaded/image/2022-3/2022-3-dop/9.jpg).

Результаты проведенного анализа

В рамках разработанной автоматизированной системы были получены данные по всем загруженным в БД КТ и КС, отдельно проанализированы русскоязычные и англоязычные КТ для тематических направлений математика, физика, информатика. Результаты анализа приведены в таблице 1. График (см. http://www.swsys.ru/uploaded/image/2022-3/2022-3-dop/10.jpg) демонстрирует распределение частоты встречаемости русских КТ в математических журналах.

Результаты анализа русскоязычных и англоязычных КС для различных тематических направлений приведены в таблице 2. На рисунке (см. http://www.swsys.ru/uploaded/image/2022-3/2022-3-dop/11.jpg) показан график распределения частоты встречаемости русских КС в математических журналах.

В списке КТ, выделенных в русскоязычных журналах по информатике (его фрагмент приведен в таблице (см. http://www.swsys.ru/uploaded/image/2022-3/2022-3-dop/12.jpg), термин «параллельные алгоритмы» встречается 29 раз, «параллельный алгоритм» – 22 раза. Аналогично в списке англоязычных терминов по информатике КТ parallel algorithms встречается 23 раза, a parallel algorithm – 21 раз.

Если исключить из рассмотрения такие общие понятия, как алгоритм, вычисления, компьютеры и т.п., то, анализируя ядро перечня КТ по информатике, можно сделать вывод, что наиболее актуальные проблемы связаны со следующими направлениями:

– моделирование (в списке из 200 наиболее часто используемых КТ термины, связанные с этой проблемой, встречаются 436 раз:

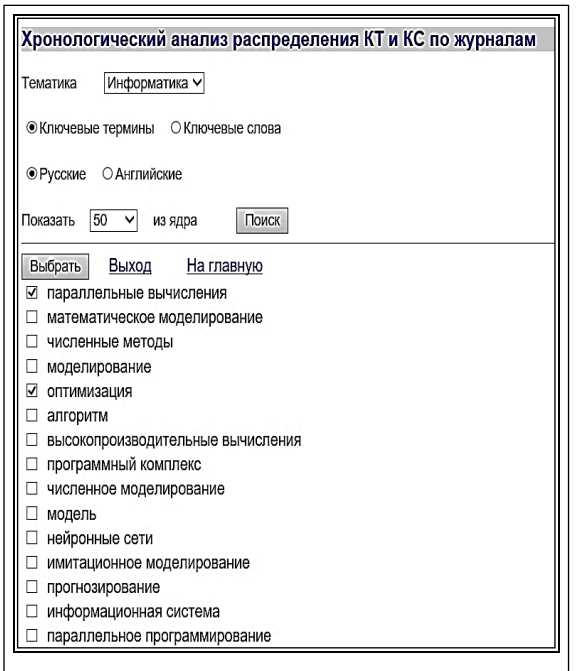

Рис. 3. Пример запроса на поиск русских КТ

Fig. 3. An example of a query to search for Russian key terms

Таблица 1

Анализ КТ

Table 1

The analysis of key terms

КТ	Математика		Физика		Информатика	
	рус.	англ.	рус.	англ.	рус.	англ.
Всего	15 752	22 969	14 095	14 013	19 031	16 684
Различных	9 928	13 533	8 047	7 975	11 230	9 535
20 % из них (наиболее повторяющихся)	1 985	2 706	1 609	1 595	2 246	1 907
Всего для выбранных 20 % (с повторениями)	7 281	11 377	6 868	6 889	9 813	8 476
Остальные	8 471	11 592	7 227	7 124	9 218	8 208
Процент повторяющихся из 20 % от всех	46,2	49,5	48,7	49,2	51,6	50,8

Таблица 2

Анализ КС

Table 2

The analysis of keywords

КС	Математика		Физика		Информатика	
	рус.	англ.	рус.	англ.	рус.	англ.
Всего	35 422	49 514	31 021	31 535	40 856	36 325
Различных	6 900	5 610	6 296	4 132	9 220	5 315
20 % из них (наиболее повторяющихся)	1 380	1 122	1 259	8 26	1 844	1 063
Всего для выбранных 20 % (с повторениями)	27 248	41 128	23 518	25 243	30 682	28 862
Остальные	8 174	8 386	7 503	6 292	10 174	7 463
Процент повторяющихся КС из 20 % от всех	76,9	83	75,8	80	75,1	79,5

математическое моделирование – 104 раза, моделирование – 96 раз, численное моделирование – 46 раз, модель – 45 раз, имитационное моделирование – 44 раза, математическая модель – 32 раза, компьютерное моделирование – 22 раза, модель данных – 10 раз, суперкомпьютерное моделирование – 10 раз, информационная модель – 9 раз, имитационная модель – 9 раз, аналитическое моделирование – 9 раз);

– параллельные вычисления (в списке из 200 наиболее часто используемых КТ термины, связанные с этой проблемой, встречаются 221 раз: параллельные вычисления – 132 раза, параллельные алгоритмы – 51 раз, параллельное программирование – 38 раз);

– оптимизация (в списке из 200 наиболее часто используемых КТ термины, связанные с этой проблемой, встречаются 68 раз: оптимизация – 50 раз, глобальная оптимизация – 10 раз, многокритериальная оптимизация – 8 раз).

Заключение

Результаты анализа показывают, что распределение КТ в том виде, как они представлены авторами, достаточно далеко от распределения Парето, в то время как распределение КС вполне ему соответствует. Более подробный анализ рейтингового списка КТ объясняет причину этого, которая в значительной степени обусловлена разной последовательностью одних и тех же слов, входящих в состав КТ.

Очевидно, что для более точной картины при обработке КТ необходимо применять методы лингвистического анализа, что на данном этапе в задачу авторов не входило. Однако сформированная БД и простой «ручной» анализ полученного ядра КТ позволяет сформировать список наиболее значимых терминов для их последующего включения в Энциклопедию и ЕЦПНЗ.

Работа выполнена в МСЦ РАН в рамках государственного задания по теме FNEF-2022-0014.

Литература

1. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестн. Российской академии наук. 2019. Т. 89. № 7. С. 728–735. DOI: 10.31857/S0869-5873897728-735.

2. Савин Г.И. Единое цифровое пространство научных знаний: цели и задачи // Информационные ресурсы России. 2020. № 5. С. 3–5. DOI: 10.51218/0204-3653-2020-5-3-5.

3. Антопольский А.Б., Белоозеров В.Н., Каленов Н.Е., Маркарова Т.С. О развитии терминологической базы данных в виде комплекса отраслевых информационно-поисковых тезаурусов // Информационные ресурсы России. 2018. № 5. С. 22–30.

4. Антопольский А.Б., Белоозеров В.Н., Маркарова Т.С. О разработке онтологии на основе классификаторов научной информации и терминологических словарей // Информационные ресурсы России. 2017. № 5. С. 2–7.

5. Каленов Н.Е. Об одном подходе к формированию предметных онтологий различных областей науки // Научный сервис в сети Интернет: тр. XXII Всерос. науч. конф. 2020. С. 276–285. DOI: 10.20948/abrau-2020-14.

6. Каленов Н.Е. Технология наполнения предметных онтологий пространства научных знаний // Электронные библиотеки. 2021. Т. 24. № 1. С. 101–115. DOI: 10.26907/1562-5419-2021-24-1-100-115.

7. Светлов Н.Н. Алгоритм поиска ключевых терминов ВЕБ-страницы // Теория и практика современной науки. 2021. № 6. С. 438–440.

8. Гринева М., Гринев М. Анализ текстовых документов для извлечения тематически сгруппированных ключевых терминов // Тр. ИСП РАН. 2009. Т. 16. С. 155–165.

9. Барахнин В.Б., Ткачев Д.А. Классификация математических документов с использованием составных ключевых терминов // ЗОНТ: матер. Междунар. конф. 2009. С. 16–23.

10. Осипова И.А. Моделирование понятийного потенциала термина «Ключевые слова» // Вестн. ОГУ. 2010. № 11. С. 117–122.

11. Никулина М.А. К вопросу об истории становления ключевых терминов в сфере информационных технологий // Актуальные проблемы гуманитарных и естественных наук. 2014. № 3-2. С. 57–60.

12. Попова Л.А. Ключевые слова современности: проблема термина // Ученые записки Петрозаводского гос. ун-та. 2017. № 5. С. 93–97.

13. Писарев И.И. Сопоставительный анализ использования ключевых англоязычных и русскоязычных терминов политики групп интересов // Вестн. МГИМО-Университета. 2018. № 6. С. 212–241. DOI: 10.24833/2071-8160-2018-6-63-212-241.

14. Квасова М.А. Сравнительный анализ ключевых русских и французских терминов в области информационных технологий // Вестн. МГЛУ. 2014. № 9. С. 64–76.

15. Koch R. The 80/20 Principle: The Secret of Achieving More with Less. London, Nicholas Brealey publ., 1998, 302 p.

Software & Systems
DOI: 10.15827/0236-235X.139.420-427

Received 08.06.22
2022, vol. 35, no. 3, pp. 420–427

An automated system for key terms analysis

S.A. Vlasova [1], *Ph.D. (Engineering), Leading Researcher, svlasova@jscc.ru*
N.E. Kalenov [1], *Dr.Sc. (Engineering), Professor, Chief Researcher, nkalenov@jscc.ru*

[1] *Joint Supercomputer Center of the Russian Academy of Sciences – JSCC, Moscow, 119334, Russian Federation*

Abstract. The paper describes the method proposed by the authors for forming an array of key terms that form the basis of subject ontology of the Common Digital Space of Scientific Knowledge and encyclopedias list of articles (slots) in a certain scientific direction.

The method is based on a frequency analysis of the occurrence of key terms in the articles published in leading scientific journals on this topic. The technique involves program processing of metadata of articles reflected in various databases, constructing rating lists of the frequency of occurrence of individual key terms and selecting the cores of such lists, which, in turn, can be considered as the basis for filling a subject ontology and forming encyclopedic slots.

To implement the methodology, the authors developed the corresponding database structure, software tools for filling it, processing and analyzing data. The paper presents the database description and the results of the practical implementation of the methodology based on processing several thousand articles from leading Russian journals in mathematics, informatics and physics (terms in Russian and English were identified and analyzed).

An assessment of the correspondence of key terms distribution frequencies and their constituent individual words to the Bradford law has shown significant discrepancies with this law in the case of key terms, but there was certain convergence when considering individual words and their permutation within key terms.

Keywords: key terms, article metadata, frequency analysis, database, automated system.

Acknowledgements. *The work was carried out at the JSC RAS within the framework of the state assignment on the topic FNEF-2022-0014.*

References

1. Antopolskii A.B., Kalenov N.E., Serebryakov V.A., Sotnikov A.N. Common digital space of scientific knowledge. *Vestn. Rossijskoj Akademii Nauk*, 2019, vol. 89, no. 7, pp. 728–735. DOI: 10.31857/S0869-5873897728-735 (in Russ.).
2. Savin G.I. Common digital space of scientific knowledge: Goals and tasks. *Information Resources of Russia*, 2020, no. 5, pp. 3–5. DOI: 10.51218/0204-3653-2020-5-3-5 (in Russ.).
3. Antopolskii A.B., Beloozerov V.N., Kalenov N.E., Markarova T.S. About development of terminological database in the form of a complex of branch information retrieval thesauruses. *Information Resources of Russia*, 2018, no. 5, pp. 22–30 (in Russ.).
4. Antopolskii A.B., Beloozerov V.N., Markarova T.S. On the development of ontology on the basis of classifiers of scientific information and terminological dictionaries. *Information Resources of Russia*, 2017, no. 5, pp. 2–7 (in Russ.).
5. Kalenov N.E. About one approach to the formation of subject ontologies for science various fields. *Proc. Sci. Conf. Scientific Service on the Internet*, 2020, pp. 276–285. DOI: 10.20948/abrau-2020-14 (in Russ.).
6. Kalenov N.E. Technology for filling subject ontologies of the scientific knowledge space. *RDLJ*, 2021, vol. 24, no. 1, pp. 101–115. DOI: 10.26907/1562-5419-2021-24-1-100-115 (in Russ.).
7. Svetlov N.N. Search algorithm for web-page keywords. *Theory and Practice of Modern Science*, 2021, no. 6, pp. 438–440 (in Russ.).
8. Grineva M., Grinev M. Parsing text documents to extract thematically grouped key terms. *Proc. of ISP RAS*, 2009, vol. 16, pp. 155–165 (in Russ.).
9. Barakhnin V.B., Tkachev D.A. Classification of mathematical documents using compound key terms. *Proc. Int. Conf. KONT*, 2009, pp. 16–23 (in Russ.).
10. Osipova I.A. Modeling the conceptual potential of the term Keywords. *Vestn. OSU*, 2010, no. 11, pp. 117–122 (in Russ.).
11. Nikulina M.A. On the issue of the history of key terms formation in the field of information technology. *Relevant Problems of the Humanities and Natural Sciences*, 2014, no. 3-2, pp. 57–60 (in Russ.).
12. Popova L.A. Key words of modern times: the issue of the term. *Proceedings of Petrozavodsk State University*, 2017, no. 5, pp. 93–97 (in Russ.).
13. Pisarev I.I. Comparative analysis of the key English and Russian language terms in studies of interest group politics. *MGIMO Review of International Relations*, 2018, no. 6, pp. 212–241. DOI: 10.24833/2071-8160-2018-6-63-212-241 (in Russ.).
14. Kvasova M.A. Comparative analysis of Russian and French key IT terms. *Vestn. of MSLU*, 2014, no. 9, pp. 64–76 (in Russ.).
15. Koch R. *The 80/20 Principle: The Secret of Achieving More with Less*. London, Nicholas Brealey Publ., 1998, 302 p.

Для цитирования

Власова С.А., Каленов Н.Е. Автоматизированная система анализа ключевых терминов // Программные продукты и системы. 2022. Т. 35. № 3. С. 420–427. DOI: 10.15827/0236-235X. 139.420-427.

For citation

Vlasova S.A., Kalenov N.E. An automated system for key terms analysis. *Software & Systems*, 2022, vol. 35, no. 3, pp. 420–427 (in Russ.). DOI: 10.15827/0236-235X.139.420-427.