

УДК 004.912

DOI: 10.25559/SITITO.15.201904.924-931

Разработка программного компонента для выявления методом словообразования хронологического порядка возникновения терминов

И. Н. Полякова*, Е. А. Филимонова

Московский государственный университет имени М.В. Ломоносова, г. Москва, Россия
119991, Россия, г. Москва, Ленинские горы, д. 1

* polyakova@cs.msu.ru

Аннотация

Родство слов является интереснейшей проблемой лингвистики русского языка, решить которую не так просто. Связи между родственными словами не всегда четко прослеживаются в связи с изменениями в языке. И близкие и схожие по происхождению слова становятся совсем непохожи друг на друга. Автоматически понять как связаны два слова - нетривиальная задача. Для реализации задачи поиска хронологического порядка возникновения терминов требуются методы, позволяющие по двум заданным словам определять последовательность появления их друг относительно друга. В предлагаемой работе ставится задача разработки универсальных методов выявления хронологического порядка возникновения слов. Выделено три основных метода - метод словообразования, метод этимологических словарей, метод гипонимов и гиперонимов. Основное внимание уделено методу словообразования, как одному из основных для решения поставленной задачи. Основа метода - сравнение морфемного строения заданных слов. По способу словообразования соответствующий метод можно разделить на несколько способов применительно к поставленной задаче: приставочный способ, суффиксальный способ, приставочно-суффиксальный способ, бессуффиксный способ и способ слияния.

Программный компонент реализован таким образом, что для двух слов на входе можно узнать, каким методом одно слово образовано от другого. При определении конкретного способа словообразования используется различие в морфемном составе исследуемых слов. Лучшие результаты система показывает для бессуффиксного способа. Для анализа точности системы была подготовлена выборка, на ней проведена оценка точности системы.

Таким образом, предложено три метода, позволяющих решить задачу ранжирования слов по времени их появления и выявления хронологического порядка их возникновения. Один из методов - метод словообразования - реализован на практике и показывает хороший результат на собранной тестовой выборке.

Ключевые слова: метод словообразования, морфемный анализ, программный компонент, терминоведение, хронологический порядок.

Финансирование: Данная работа выполнена в рамках НИР «Математическое и программное обеспечение перспективных систем обработки символьной информации», проводимой на кафедре алгоритмических языков факультета вычислительной математики и кибернетики Московского государственного университета имени М.В. Ломоносова.

Для цитирования: Полякова И. Н., Филимонова Е. А. Разработка программного компонента для выявления методом словообразования хронологического порядка возникновения терминов // Современные информационные технологии и ИТ-образование. 2019. Т. 15, № 4. С. 924-931. DOI: 10.25559/SITITO.15.201904.924-931

© Полякова И. Н., Филимонова Е. А., 2019



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Development of a Software Component to Identify the Chronological Order of the Terms by the Method of Word Formation

I. N. Polyakova*, E. A. Filimonova

Lomonosov Moscow State University, Moscow, Russia

1 Leninskie Gory, Moscow 119991, Russia

*polyakova@cs.msu.ru

Abstract

The relationship of words is an interesting problem of linguistics of the Russian language, which is not so easy to solve. The relationship between related words is not always clear due to changes in language. And close and similar in origin words become quite unlike each other. Automatically understand how two words are connected-a non-trivial task.

To implement the task of finding the chronological order of occurrence of terms requires methods allowing two given words to determine the sequence of their appearance relative to each other. The proposed work aims at developing universal methods for identifying the chronological order in which words occur. There are three main methods-the method of word formation, the method of etymological dictionaries, the method of hyponyms and hyperonyms. The main attention is paid to the method of word formation, as one of the main for solving the problem. The basis of the method is a comparison of the morphemic structure of given words. According to the method of word formation, the corresponding method can be divided into several ways in relation to the task: the prefix method, the suffix method, the prefix-suffix method, the suffixless method and the fusion method.

The software component is implemented in such a way that for two words on the input you can find out how one word is formed from another. In determining the specific method of word formation, the difference in the morphemic composition of the studied words is used. The system shows the best results for the suffixless method. To analyze the accuracy of the system, a sample was prepared and the accuracy of the system was evaluated.

Thus, three methods are proposed to solve the problem of ranking words by the time of their appearance and identifying the chronological order of their occurrence. One of the methods - the method of word formation-is implemented in practice and shows a good result on the collected test sample.

Keywords: method of word formation, morphemic analysis, program component, chronological order, terminology.

Funding: This work was carried out as part of the research work "Mathematical and Software Support of Advanced Symbolic Information Processing Systems" conducted in the Department of Algorithmic Languages of the Faculty of Computational Mathematics and Cybernetics of the M.V. Lomonosov Moscow State University.

For citation: Polyakova I.N., Filimonova E.A. Development of a Software Component to Identify the Chronological Order of the Terms by the Method of Word Formation. *Sovremennye informacionnye tehnologii i IT-obrazovanie* = Modern Information Technologies and IT-Education. 2019; 15(4):924-931. DOI: 10.25559/SITITO.15.201904.924-931



Введение

Известно, что в русском языке между родственными словами существуют связи. И эти связи не всегда четко прослеживаются. Родство слов является интереснейшей проблемой лингвистики русского языка, решить которую оказывается не так просто. С течением времени слова приобретают новые формы. Добавляя к словам, например, приставки, люди получают новые слова, которые могут иметь совсем другой смысл и лишь отдаленно по звучанию напоминать о своем предке. Близкие и схожие по происхождению слова становятся часто совсем непохожи друг на друга [1, 2]. В исследовательских задачах прикладной лингвистики часто необходимо ранжировать термины по времени их появления. Порядок возникновения понятий нужен и в задачах по терминоведению. Автоматически при анализе текста понять, как связаны два слова - нетривиальная задача.

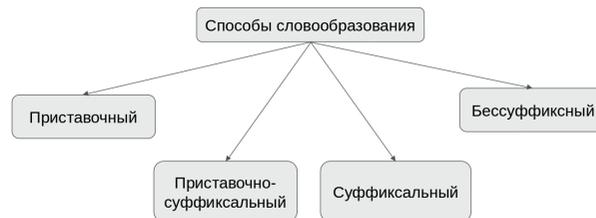
В предлагаемой работе ставится задача разработки универсальных методов выявления хронологического порядка возникновения терминов [1, 3]. Основное внимание уделено методу словообразования, как одному из основных методов решения поставленной задачи.

Предлагаемые методы

Для реализации задачи поиска хронологического порядка возникновения терминов требуются методы, позволяющие по двум заданным терминам определять последовательность появления их друг относительно друга [4, 5]. Можно выделить три основных метода - метод словообразования, метод этимологических словарей метод гипонимов и гиперонимов [6].

Первый метод основан на сравнении морфемного строения заданных терминов (слов) [7, 8]. По способу словообразования соответствующий метод можно разделить на несколько способов применительно к поставленной задаче: приставочный способ, суффиксальный способ, приставочно-суффиксальный способ, бессуффиксный способ и способ слияния [9]. Приставочный способ словообразования формирует новое слово путем добавления приставки к основе. Суффиксальный способ словообразования формирует новые слова путем добавления суффикса к основе. Приставочно-суффиксальный способ основан на двух описанных выше способах словообразования. Бессуффиксный способ формирует новые слова при помощи нулевого суффикса. Способ слияния формирует новые слова сложением уже существующих слов.

Практическую основу для реализации первого метода составляет программный инструмент, позволяющий с высокой точностью разделять входное слово по морфемам. Он не ограничен рамками той или иной предметной области, поэтому полученная морфемная структура термина [10] позволяет применить правила словообразования в русском языке, у которых наибольший интерес представляют: приставочный, суффиксальный, приставочно-суффиксальный, бессуффиксный. Также к вышеперечисленным способам можно добавить еще постфиксный способ.



Р и с. 1. Способы словообразования

Fig. 1. Word formation methods

Приставочный способ словообразования формирует новое слово добавлением приставки к основе. Стоит отметить, что приставочный способ словообразования не образует новой части речи. Например, слово “демобилизация” образовалось от слова “мобилизация” приставочным способом, или слово “предсказание” образовано от “сказание”. Схема работы приставочного способа применительно к частям речи выглядит следующим образом:

- Глаг. -> Глаг.
- Сущ. -> Сущ.
- Прил. -> Прил.
- Местоим. -> Местоим.
- Нар. -> Нар.

Суффиксальный способ словообразования формирует новые слова добавлением суффикса к основе. Этот способ отличается от предыдущего и тем, что при его использовании могут образовываться слова, которые будут иметь часть речи, отличную от части речи первого слова. Например, слово “рыночная” образовалось от слова “рынок” суффиксальным способом. Таким образом, схема работы суффиксального способа применительно к частям речи выглядит следующим образом:

- Глаг. -> Глаг., Сущ., Прил.
- Сущ. -> Сущ., Прил., Глаг.
- Прил. -> Глаг., Нар., Сущ.

Приставочно-суффиксальный способ основан на одновременном применении приставочного и суффиксального способов. Например, от слова “оружие” приставочно-суффиксальным способом образуется слово “обезоруженный”. Схема работы приставочно-суффиксального способа применительно к частям речи выглядит следующим образом:

- Сущ. -> Сущ., Прил., Глаг., Нар.
- Прил. -> Глаг., Нар.
- Числ. -> Нар.

Бессуффиксный способ формирует новые слова при помощи нулевого суффикса. При этом нулевой суффикс в письме и в речи никак не выражается. Бессуффиксный способ позволяет менять часть речи. Например, от слова “гладкий” и образовалось слово “гладь”.

- Глаг. -> Сущ.
- абстрактное значение действия(подскочить-подскок, бежать-бег)
- предметное значение, значение лица, производящего действие(всходить-всход)
- значение места действия(подрывать-подрыв, заливать-залив)
- значение объекта или результата действия(взрывать-взрыв)
- Прил. -> Сущ.
- отвлеченное значение признака(гладкий-гладь)



- значение носителя признака(интеллигентный-интеллигент)
- значение человека,выполняющего действие(обслуживать-обслуга)
- Сущ. -> Прил.
- с общим значением(золото-золотой)
- отсутствие(без хвоста-безхвостый)
- принадлежности(отец-отчий)
- Глаг. -> Прил.
- значение признака по действию (заехать-заезжий)
- Нар. -> Прил.
- значение превосходящей степени признака(ниже-нижший)
- Числ.
- от количественного в качественное(пять-пятый)

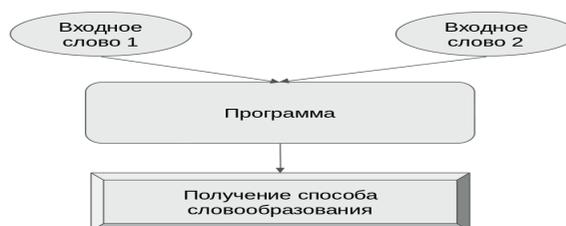
Бессуффиксный способ, как и приставочный, приставочно-суффиксальный и суффиксальный, весьма удобен для решения поставленной задачи, однако не решает ее полностью. Безусловно, при использовании вышеперечисленных способов в первом методе может возникнуть неоднозначность. Неточность может быть разрешена в следующем методе [11]. Рассмотрим его обзорно.

Метод выявления хронологического порядка слов на основе использования этимологических словарей. В настоящее время существует большое количество словарей, в которых по слову можно найти слово-прародителя, от которого образовалось первое. Это слово-прародитель прописано явно, и, найдя его, мы получаем правильный результат. Неоднозначностей метод не порождает, однако этот метод не может быть применен, если слово не описано в словаре. Так, например, для слова “политика” в этимологических словарях можно найти слово, от которого оно образовано - это слово “город”. Или, допустим, другой пример, где слово “общество” произошло от слова “общий”. Эта информация дает полный и безошибочный результат, поскольку информация в этимологических словарях является достоверной [12-14]. И, наконец, рассмотрим третий метод образования слов.

Метод выявления порядка слов на основе выделения обобщения и частного, более известный как задача нахождения гипонимов и гиперонимов [15]. Гипоним — понятие, выражающее частную сущность по отношению к другому, более общему понятию. Гипероним — слово с более широким значением, выражающее общее, родовое понятие, название класса (множества) предметов (свойств, признаков). Этот метод позволит решать задачу для случаев, когда более широкое понятие образовалось путем объединения частных, и, наоборот. Если один из терминов является гипонимом, а другой - гиперонимом, и при этом у них есть общее слово, то гипероним появился раньше гипонима. Например, для пары понятий - “искусство” и “театральное искусство” - понятие “искусство” является гиперонимом, а понятие “театральное искусство” - гипонимом, откуда следует, что понятие “театральное искусство” появилось позже понятия “искусство” [16, 17, 18]. Сложность представляют случаи, когда гипоним и гипероним синтаксически не схожи. В этом случае определить, какое именно понятие возникло раньше без каких-либо других пособий невозможно. Поэтому данный метод не всегда выдает верный результат.

Программная реализация

Программный инструмент реализован таким образом, что для двух слов на входе она позволяет узнать каким методом одно слово образовано от другого.

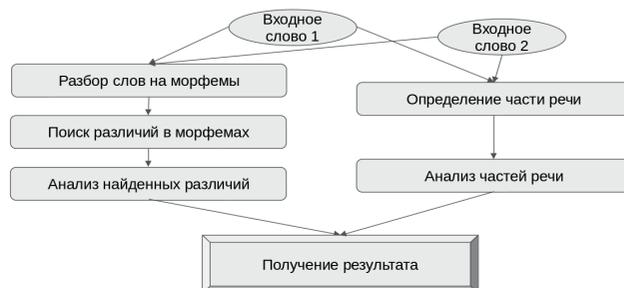


Р и с. 2. Общая схема работы программы

F i g. 2. The general scheme of the program

Для определения способа образования необходима информация о морфемном строении слов и о частях речи слов. Для морфемного анализа используются готовые системы - рассматривались наиболее популярные сайты с онлайн-системами и результаты брались оттуда. Были выбраны и протестированы лучшие онлайн-системы, и выбраны те, которые показали наилучший результат. Разумеется, разделение на части речи с помощью этих сайтов было не всегда точным и возможным, однако точность достаточно высока.

Для определения части речи используется анализатор рунотгрфу, который позволил определять часть слова с очень высокой точностью, однако встречаются такие слова, для которых нельзя однозначно определить часть речи без какого-либо другого контекста [19, 20, 21, 22]. При поиске конкретного способа словообразования в реализации программы применяется идея о поиске различий в разборе слов [23].



Р и с. 3. Подробная схема работы программы

F i g. 3. Detailed scheme of the program

При этом важно рассматривать случай чередования корней, в противном случае программа не видела бы в словах сходства, потому что слова с разными корнями не являются словами, которые образовались одно от другого приставочным, суффиксальным, приставочно-суффиксальным или бессуффиксным способом. Примером такого чередования корней является

[('change', ['корень', 0], ('рек', 'реч')), ('add', ',', [['суффикс', ['н']]])]
река -> речной : суффиксальный

Для проверки работоспособности метода использована специально подготовленная тестовая выборка.



Таблица 1. Состав тестовой выборки
Table 1. Test sampling pattern

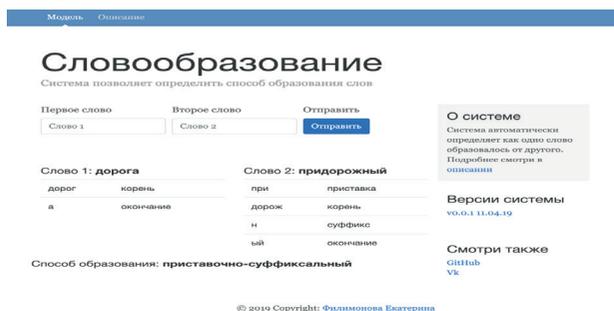
Способ словообразования	Количество примеров
Бессуффиксный	230
Суффиксальный	98
Приставочный	82
Приставочно-суффиксальный	58

Точность работы метода представлена в таблице 2. Лучшие результаты система показывает для бессуффиксного способа словообразования слов. Анализ ошибок, допускаемых системой, показывает, что почти все они связаны с неправильным выделением морфем слов. Чем лучше работает система, которая выделяет морфемы слов, тем выше точность предложенной системы.

Таблица 2. Точность метода словообразования на тестовой выборке
Table 2. Accuracy of the word formation method on a test sample

Способ образования	Точность, %
Бессуффиксный	84
Суффиксальный	80
Приставочный	76
Приставочно-суффиксальный	71
Общая точность	79

Для наглядности работы программы был реализован сайт (рис. 4) [24, 25], где можно увидеть то, как работает программа — ее входные данные и результат. Пример работы программы:



Р и с. 4. Работы сайта, демонстрирующего работу программы
F i g. 4. Website demonstrating the work of the program

Заключение

В данной работе описана задача разработки универсальных методов выявления хронологического порядка возникновения терминов. В качестве возможных решений задачи предложены три метода, которые позволяют решить задачу ранжирования слов по времени их появления и выявления хронологического порядка возникновения. Один из методов - метод словообразования - реализован на практике. В реализации метода применяется поиск различий в морфемном разборе слов. Разбор на морфемы берется из уже существующих систем и производится анализ полученной информации. Для анализа также используется информация о частях речи каждого из слов.

Была собрана тестовая выборка для проверки качества работы системы. Выборка представляет собой множество из пар слов, которое сбалансировано по классам словообразования [26, 27]. На этой выборке проведены эксперименты и получены результаты. Лучшее качество система показывает для бессуффиксного способа словообразования слов.

Тем не менее реализована лишь часть методов, которые предложены для решения задачи. Словообразование играет одну из важнейших ролей в определении хронологического порядка появления и образования слов. Для более точного результата работы предложены еще два метода. Их тоже планируется в дальнейшем реализовать для решения поставленной задачи [28, 29, 30].

Список использованных источников

- [1] Мальковский М. Г., Соловьев С. Ю. Универсальное терминологическое пространство // Труды международного семинара «Компьютерная лингвистика и интеллектуальные технологии». Т. 1. М.: Наука, 2002. С. 266-270. URL: <http://park.glossary.ru/serios/theory01.php> (дата обращения: 28.10.2019).
- [2] Соловьев С. Ю. Схема и формула глоссария // Сб. Трудов XXI национальной конференции по искусственному интеллекту с международным участием. Т. 2. М.: ЛЕНАНД, 2008. С. 157-164. URL: <https://park.glossary.ru/serios/theory08.php> (дата обращения: 28.10.2019).
- [3] Соловьев С. Ю. Образные представления терминологической сети // Прикладное программное обеспечение. М.: Изд-во МИРЭА, 2008. С. 55-69. URL: <http://park.glossary.ru/serios/theory06.php> (дата обращения: 28.10.2019).
- [4] Loukachevitch N. V., Dobrov B. V. The Sociopolitical Thesaurus as a resource for automatic document processing in Russian // Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication. 2015. Vol. 21, Issue 2. Pp. 237-262. DOI: 10.1075/term.21.2.05lou
- [5] Мальковский М. Г., Соловьев С. Ю. Методы формирования глоссариев в универсальном терминологическом пространстве // Труды международной конференции «Компьютерная лингвистика и интеллектуальные технологии». М.: Наука, 2003. С. 438-440. URL: <https://www.park.glossary.ru/serios/theory02.php> (дата обращения: 28.10.2019).
- [6] Филимонова Е. А., Соловьев С. Ю., Полякова И. Н. Разработка универсальных методов выявления хронологического или псевдохронологического порядка возникновения терминов в заданной предметной области / Под ред. В. В. Голенкова // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2019). Материалы международной научно-технической конференции, Минск, 21 - 23 февраля 2019 г. Минск: БГУИР, 2019. С. 285-288. URL: <https://libelodoc.bsuir.by/handle/123456789/34571> (дата обращения: 28.10.2019).
- [7] Лейчик В. М. Терминоведение: Предмет, методы, структура. М.: Книжный дом «ЛИБРОКОМ», 2009. 256 с.
- [8] Мальковский М. Г., Соловьев С. Ю. Исследование родовых отношений в терминологических сетях // Открытые семантические технологии проектирования интеллектуальных систем. 2013. № 3. С. 147-152 URL:



- <https://www.elibrary.ru/item.asp?id=30115162> (дата обращения: 28.10.2019).
- [9] Земская Е. А. Современный русский язык. Словообразование. М.: Флинта: Наука, 2011. 328 с.
- [10] *Vandenbussche P-Y, Charlet J. Méta-modèle général de description de ressources terminologiques et ontologiques // IC 2009 - 20èmes Journées Francophones d'Ingénierie des Connaissances, May 2009, Hammamet, Tunisie. à paraître. fihal-00379935f URL: <https://hal.archives-ouvertes.fr/hal-00379935> (дата обращения: 28.10.2019).*
- [11] Гринева-Гриневиц С. В. Введение в терминографию. М.: ЛИБРОКОМ, 2009. 224 с.
- [12] *Kalyanpur A, Parsia B, Hendler J. A Tool for Working with Web Ontologies // International Journal on Semantic Web and Information Systems. 2005. Vol. 1, No. 1. Pp. 36-39. DOI: 10.4018/jswis.2005010103*
- [13] Можарова В. А., Лукашевич Н. В. Исследование признаков для извлечения именованных сущностей из текстов на русском языке // Научно-техническая информация. С. 2: Информационные процессы и системы. 2017. № 5. С. 14-21. URL: <https://www.elibrary.ru/item.asp?id=29149149> (дата обращения: 28.10.2019).
- [14] Нокель М. А., Лукашевич Н. В. Тематические модели в задаче извлечения однословных терминов // Программная инженерия. 2014. № 3. С. 34-40. URL: <https://www.elibrary.ru/item.asp?id=21257444> (дата обращения: 28.10.2019).
- [15] Шелов С. Д. Термин. Терминологичность. Терминологические определения. СПб.: Изд-во СПбГУ, 2003. 279 с.
- [16] Апресян Ю. Д. Избранные труды. Т. 1. Лексическая семантика, 2-е изд., испр. и доп. М.: «Восточная литература» РАН, 1995. 472 с.
- [17] *Gubanov D. A., Makarenko A. V., Novikov D. A. Analysis methods for the terminological structure of a subject area // Automation and Remote Control. 2014. Vol. 75, Issue 12. Pp. 2231-2247. DOI: 10.1134/S00051179141201331*
- [18] Лукашевич Н. В., Герасимова А. А. Определение устойчивых словосочетаний методом ассоциативного эксперимента // Вестник Московского университета. Серия 9: Филология. 2018. № 1. С. 23-42. URL: <https://www.elibrary.ru/item.asp?id=32833935> (дата обращения: 28.10.2019).
- [19] Мальковский М. Г., Соловьев С. Ю. Терминологические сети // Открытые семантические технологии проектирования интеллектуальных систем. 2012. № 2. С. 77-82. URL: <https://www.elibrary.ru/item.asp?id=32704351> (дата обращения: 28.10.2019).
- [20] *Smith K. W. Cython: A Guide for Python Programmers. O'Reilly Media, Inc., 2015. 254 pp.*
- [21] Лутц М. Изучаем Python, 5-е изд., Т. 1. М.: Диалектика, 2019. 832 с.
- [22] *Formica A. Similarity reasoning in formal concept analysis: from one- to many-valued contexts // Knowledge and Information Systems. 2019. Vol. 60, Issue 2. Pp. 715-739. DOI: 10.1007/s10115-018-1252-4*
- [23] Мальковский М. Г., Соловьев С. Ю. Правила формирования терминологических кластеров // Открытые семантические технологии проектирования интеллектуальных систем. 2014. № 4. С. 169-172. URL: <https://www.elibrary.ru/item.asp?id=30080190> (дата обращения: 28.10.2019).
- [24] *Соловьев С. Ю. Об одном методе генерации страниц-карт для веб-сайтов // Информационные процессы. 2008. Т. 8, № 1. С. 24-29. URL: <https://www.elibrary.ru/item.asp?id=17685281> (дата обращения: 28.10.2019).*
- [25] *Lutz M. Learning Python. 5th Edition. O'Reilly Media, 2013. 1648 pp.*
- [26] *Alekseev A. A., Loukachevitch N. V. The automatic retrieval of news entities based on the structure of a news cluster // Scientific and Technical Information Processing. 2012. Vol. 39, Issue 6. Pp. 303-309. DOI: 10.3103/S0147688212060019*
- [27] *Mozharova V. A., Lukashovich N. V. Investigation of features for extraction of named entities from texts in Russian // Automatic Documentation and Mathematical Linguistics. 2017. Vol. 51, Issue 3. Pp. 127-134. DOI: 10.3103/S0005105517030049*
- [28] Кузнецов С. О. Методы теории решеток и анализа формальных понятий в машинном обучении // Новости искусственного интеллекта. 2004. № 3. С. 19-31. URL: <http://www.raai.org/library/ainews/getainews.php?2004> (дата обращения: 28.10.2019).
- [29] *Ganter B., Wille R. Formal Concept Analysis: Mathematical Foundations. Berlin: Springer Science & Business Media, 1999. 284 pp. DOI: 10.1007/978-3-642-59830-2*
- [30] *Freixa J., Fernández-Silva S. Terminological variation and cognition: cognitive factors in the name of a specialized concept // Multiple Perspectives on Terminological Variation / P. Drouin, A. Francœur, J. Humbley, A. Picton (Eds.) John Benjamins Publishing Company, 2017. Pp. 155-180. DOI: 10.1075/tlrp.18.07fre*

Поступила 28.10.2019; принята к публикации 30.11.2019;
опубликована онлайн 23.12.2019.

Об авторах:

Полякова Ирина Николаевна, доцент кафедры алгоритмических языков, факультет вычислительной математики и кибернетики, Московский государственный университет имени М.В. Ломоносова (119991, Россия, г. Москва, Ленинские горы, д. 1), кандидат физико-математических наук, ORCID: <http://orcid.org/0000-0003-1432-4906>, polyakova@cs.msu.ru

Филимонова Екатерина Александровна, магистрант, кафедра алгоритмических языков, факультет вычислительной математики и кибернетики, Московский государственный университет имени М.В. Ломоносова (119991, Россия, г. Москва, Ленинские горы, д. 1), ORCID: <http://orcid.org/0000-0003-1155-4386>, filkate0903@rambler.ru

Благодарности: авторы выражают особую благодарность за поддержку, помощь и ценные замечания профессору кафедры алгоритмических языков факультета вычислительной математики и кибернетики Московского государственного университета имени М.В. Ломоносова Сергею Юрьевичу Соловьеву.

Все авторы прочитали и одобрили окончательный вариант рукописи.



References

- [1] Malkovsky M.G., Soloviev S.Y. Universal terminological space. In: *Proceedings of the international seminar "Computer linguistics and intellectual technologies"*, vol. 1. Nauka, Moscow, 2002, pp. 266-270. Available at: <http://park.glossary.ru/serios/theory01.php> (accessed 28.10.2019). (In Russ.)
- [2] Soloviev S.Y. Diagram and formula of the Glossary. In: *Proceedings of the XXI national conference on artificial intelligence with international participation*, vol. 2. LENAND, Moscow, 2008, pp. 157-164. Available at: <http://park.glossary.ru/serios/theory08.php> (accessed 28.10.2019). (In Russ.)
- [3] Soloviev S.Y. Figurative representations of the terminological network. In: *Application Software*. MIREA, Moscow, 2008, pp. 55-69. Available at: <http://park.glossary.ru/serios/theory06.php> (accessed 28.10.2019). (In Russ.)
- [4] Loukachevitch N. V., Dobrov B. V. The Sociopolitical Thesaurus as a resource for automatic document processing in Russian. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*. 2015; 21(2):237-262. (In Eng.) DOI: 10.1075/term.21.2.05lou
- [5] Malkovsky M.G., Soloviev S.Y. Methods of formation of glossaries in the universal terminological space. In: *Proceedings of the international conference "Computer linguistics and intellectual technologies"*. Nauka, Moscow, 2003, pp. 438-440. Available at: <https://www.park.glossary.ru/serios/theory02.php> (accessed 28.10.2019). (In Russ.)
- [6] Filimonova E.A., Soloviev S.Yu., Polyakova I.N. Development of universal detection methods for identifying chronological or pseudo-chronological order of occurrence of terms in a given subject area. In: *Proceedings of the international conference "Open Semantic Technologies for Intelligent Systems" (OSTIS-2019)*. BSUIR, Minsk, 2019, pp. 285-288. Available at: (accessed 28.10.2019). (In Eng. abstract in Russ.)
- [7] Lejchik V.M. *Terminovedenie: Predmet, metody, struktura* [Terminology: Subject, methods, structure]. Librokom, Moscow, 2009. (In Russ.)
- [8] Malkovsky M.G., Soloviev S.Y. Hierarchical Relations in Terminological Network. *Open Semantic Technologies for Intelligent Systems*. 2013; (3):147-152. Available at: <https://www.elibrary.ru/item.asp?id=30115162> (accessed 28.10.2019). (In Russ. abstract in Eng.)
- [9] Zemskaya E.A. *Sovremennyj russkij yazyk. Slovoobrazovanie* [Modern Russian Language: Word-formation]. Flinta: Nauka, Moscow, 2011. (In Russ.)
- [10] Vandenbussche P-Y., Charlet J. Méta-modèle général de description de ressources terminologiques et ontologiques. In: *IC 2009 - 20èmes Journées Francophones d'Ingénierie des Connaissances*, May 2009, Hammamet, Tunisie. à paraître. fhal-00379935f Available at: <https://hal.archives-ouvertes.fr/hal-00379935> (accessed 28.10.2019). (In French)
- [11] Grinev-Grinevich S.V. *Vvedenie v terminografiyu* [Introduction to terminography]. Librokom, Moscow, 2009. (In Russ.)
- [12] Kalyanpur A., Parsia B., Hendler J. A Tool for Working with Web Ontologies. *International Journal on Semantic Web and Information Systems*. 2005; 1(1):36-39. (In Eng.) DOI: 10.4018/jswis.2005010103
- [13] Mozharova V.A., Lukashevich N.V. Examination of the indications for the extraction of named entities from texts in Russian. *Scientific and Technical Information Processing. Series 2: Information Processes and Systems*. 2017; (5):14-21. Available at: <https://www.elibrary.ru/item.asp?id=29149149> (accessed 28.10.2019). (In Russ.)
- [14] Nokol M.A., Loukachevitch N.V. Topic Models in the Task of Single-Word Term Extraction. *Programmnyaya inzheneriya = Software Engineering*. 2014; (3):34-40. Available at: <https://www.elibrary.ru/item.asp?id=21257444> (accessed 28.10.2019). (In Russ., abstract in Eng.)
- [15] Shelov S.D. *Termin. Terminologichnost'. Terminologicheskie opredeleniya* [A Technical Term. Its Terminological Side. On Definitions of Technical Terms]. St. Petersburg, SPbSU, 2003. (In Russ.)
- [16] Apresyan Yu.D. *Izbrannye trudy. T. 1. Leksicheskaya semantika* [Selected Works. Vol. 1: Lexical Semantics (Synonymous Means of Language)]. 2nd ed. Vostochnaya literatura, RAS, Moscow, 1995. (In Russ.)
- [17] Gubanov D.A., Makarenko A.V., Novikov D.A. Analysis methods for the terminological structure of a subject area. *Automation and Remote Control*. 2014; 75(12):2231-2247. (In Eng.) DOI: 10.1134/S00051179141201331
- [18] Lukashevich N.V., Gerasimova A.A. Detecting Conventionalized Multiword Expressions by Using a Word Association Experiment. *Moscow State University Bulletin. Series 9. Philology*. 2018; (1):23-42. Available at: <https://www.elibrary.ru/item.asp?id=32833935> (accessed 28.10.2019). (In Russ., abstract in Eng.)
- [19] Malkovskii M.G., Soloviev S.Yu. Terminological Networks. *Open Semantic Technologies for Intelligent Systems*. 2012; (2):77-82. Available at: <https://www.elibrary.ru/item.asp?id=32704351> (accessed 28.10.2019). (In Russ., abstract in Eng.)
- [20] Smith K.W. *Cython: A Guide for Python Programmers*. O'Reilly Media, Inc., 2015. (In Eng.)
- [21] Lutz M. *Learning Python*, 5th ed., Dialectics, Moscow, 2019. (In Russ.)
- [22] Formica A. Similarity reasoning in formal concept analysis: from one- to many-valued contexts. *Knowledge and Information Systems*. 2019; 60(2):715-739. (In Eng.) DOI: 10.1007/s10115-018-1252-4
- [23] Malkovskii M.G., Soloviev S.Yu. Rules for Terminological Clusters Creations. *Open Semantic Technologies for Intelligent Systems*. 2014; (4):169-172. Available at: <https://www.elibrary.ru/item.asp?id=30080190> (accessed 28.10.2019). (In Russ., abstract in Eng.)
- [24] Soloviev S.Yu. About one method of generation of pages-maps for websites. *Information Processes*. 2008; 8(1):24-29. Available at: <https://www.elibrary.ru/item.asp?id=17685281> (accessed 28.10.2019). (In Russ.)
- [25] Lutz M. *Learning Python*. 5th Edition. O'Reilly Media, 2013. (In Eng.)
- [26] Alekseev A.A., Loukachevitch N.V. The automatic retrieval of news entities based on the structure of a news cluster. *Scientific and Technical Information Processing*. 2012; 39(6):303-309. (In Eng.) DOI: 10.3103/S0147688212060019
- [27] Mozharova V.A., Lukashevich N.V. Investigation of features for extraction of named entities from texts in Russian. *Automatic Documentation and Mathematical Linguistics*. 2017; 51(3):127-134. (In Eng.) DOI: 10.3103/S0005105517030049
- [28] Kuznetsov S.O. Methods of Lattice Theory and Formal Concept Analysis in Machine Learning. *Novosti iskusstvennogo intellekta = Artificial Intelligence News*. 2004; (3):19-31. Available at: <http://www.raai.org/library/ainews/getainews.php?2004> (accessed 28.10.2019). (In Russ., abstract in Eng.)
- [29] Ganter B., Wille R. *Formal Concept Analysis: Mathematical Foundations*. Berlin: Springer Science & Business Media, 1999. (In Eng.) DOI: 10.1007/978-3-642-59830-2



- [30] Freixa J., Fernández-Silva S. Terminological variation and cognition: cognitive factors in the name of a specialized concept. In: Drouin P., Francœur A., Humbley J., Picton A. (Eds.) *Multiple Perspectives on Terminological Variation*. John Benjamins Publishing Company, 2017, pp. 155-180. (In Eng.) DOI: 10.1075/tlrp.18.07fre

*Submitted 28.10.2019; revised 30.11.2019;
published online 23.12.2019.*

About the authors:

Irina N. Polyakova, Associate Professor of the Department of Algorithmic Languages, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University (1 Leninskie Gory, Moscow 119991, Russia), Ph.D. (Phys.-Math.), ORCID: <http://orcid.org/0000-0003-1432-4906>, polyakova@cs.msu.ru

Ekaterina A. Filimonova, Master's Degree student, Department of Algorithmic Languages, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University (1 Leninskie Gory, Moscow 119991, Russia), ORCID: <http://orcid.org/0000-0003-1155-4386>, filkate0903@rambler.ru

Acknowledgments: The authors are particularly thankful to Sergei Yuryevich Soloviev, Professor of the Department of Algorithmic Languages, Faculty of Computational Mathematics and Cybernetics, Moscow State University named after MV Lomonosov for his support, help, and valuable comments.

All authors have read and approved the final manuscript.

