

**УДК 004.4**

**ББК 3**

**Елизаров А.М., Жильцов Н.Г., Иванов В.В., Кириллович А.В.,  
Липачёв Е.К., Невзорова О.А.**

Казанский (Приволжский) федеральный университет  
Казань, Россия

<mailto:amelizarov@gmail.com>, [nikita.zhiltsov@gmail.com](mailto:nikita.zhiltsov@gmail.com),  
[nomemm@gmail.com](mailto:nomemm@gmail.com), [aлик.kirillovich@gmail.com](mailto:aлик.kirillovich@gmail.com), [elipachev@gmail.com](mailto:elipachev@gmail.com),  
[onevzoro@gmail.com](mailto:onevzoro@gmail.com)

## **СЕМАНТИЧЕСКИЙ РЕКОМЕНДАТЕЛЬНЫЙ СЕРВИС В ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ МАТЕМАТИКА**

**Аннотация:** Предложен комплекс технологий, позволяющий на основе персонального профиля учёного эффективнее работать с научно-техническим контентом, а при анализе документов – выделять понятия, наиболее важные с точки зрения личного семантического профиля, и на их основе формировать связное семантическое представление документа. Разработан рекомендательный сервис, который на базе таких представлений позволяет персонализировано подбирать документы (научные публикации, определения терминов и др.), релевантные личному семантическому профилю ученого-математика. В качестве технологической основы использованы методы терминологического аннотирования и онтологическая модель *OntoMath<sup>PRO</sup>* математической области знания.

**Ключевые слова:** рекомендательные сервисы, семантическое аннотирование, онтологии, извлечение информации, информационный поиск, системы управления информацией.

**Elizarov A.M., Zhiltsov N.G., Ivanov V.V., Kirillovich A.V.,  
Lipachev E.K., Nevzorova O.A.**

Kazan (Volga Region) Federal University  
Kazan, Russia

[mailto: amelizarov@gmail.com](mailto:amelizarov@gmail.com), [nikita.zhiltsov@gmail.com](mailto:nikita.zhiltsov@gmail.com),  
[nomemm@gmail.com](mailto:nomemm@gmail.com), [aлик.kirillovich@gmail.com](mailto:aлик.kirillovich@gmail.com), [elipachev@gmail.com](mailto:elipachev@gmail.com),  
[onevzoro@gmail.com](mailto:onevzoro@gmail.com)

## **SEMANTIC-ORIENTED RECOMMENDATION SERVICE FOR PROFESSIONAL MATHEMATICIANS**

**Abstract:** We propose a technological framework that makes access to scientific content more productive. The system extracts most important concepts from the user's semantic profile, and, then, builds an integrated semantic representation of documents. We are developing a personalized recommendation service that can find relevant documents (scientific publications, definitions of novel

*terms etc.) with respect to the user profile. The service relies on methods of terminological annotations as well as OntoMath<sup>PRO</sup>, our ontology of mathematical knowledge.*

**Keywords:** recommendation systems, semantic annotation, ontology engineering, information extraction, information retrieval, content management systems.

**Введение.** Проблема управления научной информацией, входящей в круг интересов отдельного ученого, была и остается актуальной. Еще в 1945 году в [1] на базе существовавших в то время информационных технологий была предложена модель комплекса управления персональной научной информацией. Концептуальной основой этого комплекса была система ассоциативных связей – прообраз гипертекста, формируемая пользователем при работе с научной информацией. Сегодня наиболее эффективным подходом к решению проблем управления научной информацией, в частности, математической являются семантические технологии, например, [2 – 5]. В результате обработка семантических связей, выполняемая преимущественно в автоматическом режиме, позволяет формировать личное информационное пространство ученого.



Рис. 1. Этапы работы ученого с научной информацией

Важным направлением области Семантического веба стала разработка онтологий предметных областей [6], а поиск на основе онтологий показал высокую эффективность в разных приложениях [7]. Развитие методов автоматической обработки текстов позволило решать задачи аннотирования и извлечения знаний в терминах онтологий – примером этого для научного контента служит платформа ScienceWISE (<http://sciencewise.info>), которая предоставляет учёным возможность согласовывать рабочую терминологию в рамках специализированной онтологии. Далее, используя онтологию, система автоматически аннотирует научные статьи из общедоступной электронной коллекции ArXiv (<http://arxiv.org/>) и предоставляет интерфейс для

работы с документом [8]. В то же время сервисы, имеющиеся в распоряжении ученых, не обеспечивают сопровождения всех этапов работы с научной информацией (см. рис. 1).

Цели настоящей работы – автоматизация процессов поиска публикаций по теме исследований, а также получение дополнительной информации о терминологии, используемой в анализируемых публикациях.

**Семантический рекомендательный сервис.** На рис. 2 представлена архитектура разрабатываемой системы. Процесс обработки информации выглядит следующим образом. Контент электронной коллекции семантически анализируется во внешнем сервисе семантического аннотирования Textocat [9]. Результаты аннотирования – аннотации в терминах онтологии предметной области OntoMath<sup>PRO</sup> (<http://ontomathpro.org/>) – сохраняются в базе знаний сервиса. Аналитические модули обрабатывают данные из базы знаний и формируют представление в виде интерактивной карточки публикаций и понятий, выводимых на экран пользователя. Центральным элементом системы является онтология OntoMath<sup>PRO</sup> [10, 11], разработанная при участии авторов. Она содержит описания математических понятий и связей между ними.

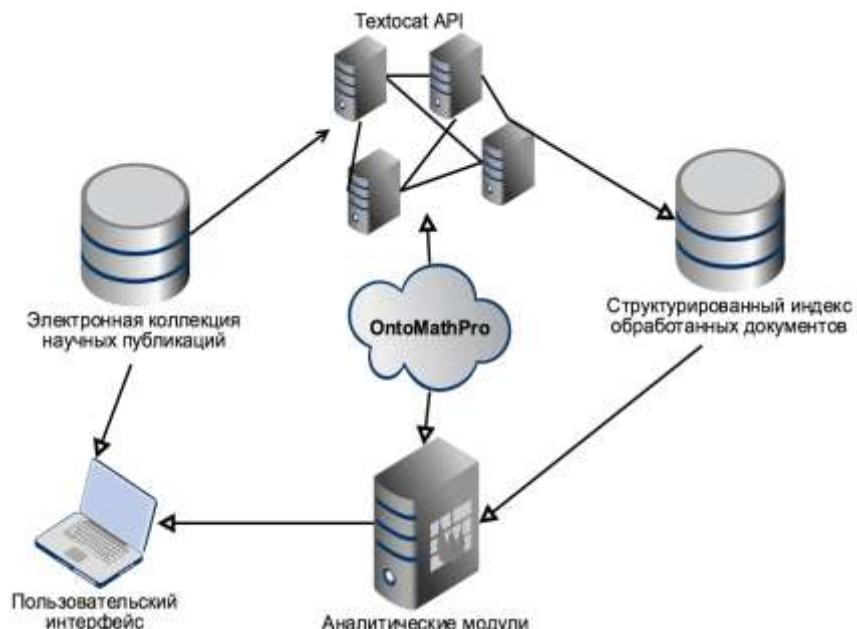


Рис. 2. Архитектура сервиса

**Модуль семантического аннотирования** решает задачу связывания фраз из описаний научных публикаций с концептами из терминологических источников. В качестве таковых разработанная система поддерживает онтологии OntoMath<sup>PRO</sup> [11] и ScienceWISE [8], а также базу знаний DBpedia [12].

На первом этапе модуль генерирует фразы – кандидаты на связывание – на основе информации, полученной средствами лексико-синтаксического анализа, реализованного в облачном сервисе текстовой аналитики Textocat

[9]. Этот анализ включает функции точного сопоставления по словарю, а также возможности выделения границ именных групп. При этом словарь составляется из имен онтологических концептов, включая синонимы, сокращения и эквивалентные понятия.

На втором этапе модуль ранжирует концепты – кандидаты на связывание – с полученными фразами. При ранжировании учитываются характеристики, выражющие строковую близость имен кандидатов, а также контекст слов и других выделенных концептов, окружающих фразу-кандидата.

На третьем этапе происходит валидация высоко ранжированных кандидатов. Каждому концепт-кандидату приписывается свое ранжирующее значение («уверенность»). В случае низкой уверенности модуля в кандидате на связывание для данной фразы связывание не производится.

**Модуль составления семантического профиля статьи** решает задачу получения структурированного графового представления документа на основе его терминологического содержимого. Онтологические концепты, выделенные в тексте научной статьи на предыдущем этапе, имеют семантические связи между собой, определенные в онтологии. При этом концепты можно трактовать как вершины некоторого семантического графа, а отношения между концептами – как ребра в графе. Тогда семантический профиль статьи – это минимальное оставное дерево, содержащее все выделенные концепты. Семантический профиль научной публикации отражает ее тематику и важнейшие понятия, без понимания которых невозможно вникнуть в суть статьи.

**Модуль составления семантического профиля пользователя** строит аналогичный семантический профиль на основе онтологических концептов, исходя из истории просмотра пользователем статей на нашем ресурсе. Профиль содержит основные концепты, с которым пользователю приходится работать наиболее часто, то есть отражает его компетенции. Это дает возможность считать, что пользователь знает определения и семантику таких терминов. Кроме того, пользователь может явно выбирать ключевые слова, которые будут соответствовать его компетенции.

**Модуль рекомендации определений** решает задачу составления рекомендаций при прочтении некоторым пользователем определенной научной публикации. Результатом являются список терминов, которые система считает наименее понятными пользователю. При этом используется информация о семантических профилях данной статьи и данного пользователя. Система может находить как термины, которые упоминаются в статье непосредственно, так и близкие понятия, которые важны, но не встречаются в тексте. С точки зрения пользовательского интерфейса составленное определение содержит текст альтернативных формулировок из разных ресурсов, а также графическое представления о связях термина с близкими понятиями из онтологии.

**Пользовательский интерфейс.** Одним из сервисов пользовательского интерфейса системы является карточка публикации (см. рис. 3 и 4), которая содержит:

- наиболее важные понятия, которые упоминаются в тексте статьи; щелкнув по ключевому слову, пользователь переходит на карточку понятия;
- рекомендованные статьи; список формируется на основе их тематической близости.

С. К. Водопьянов, И. М. Пупышев

## Следы функций из пространства Соболева на множествах Альфорса групп Карно

Сиб. матем. журн., 2007, том 48, номер 6

**Аннотация:** Доказана обратная теорема о следах функций из пространств Соболева  $W^{1,p}$ , заданных на группе Карно, на регулярных замкнутых подмножествах, называемых  $d$ -множествами Альфорса (прямая теорема о следах получена в другой работе авторов). Теорема обобщает результаты А. Йонссона и Х. Валлина для функций классов Соболева в евклидовом пространстве. В качестве следствия приводится теорема о граничных значениях функций из пространств Соболева, заданных в области с гладкой границей на двухступенчатой группе Карно. Рассматривается пример применения полученных теорем к разрешимости краевой задачи для одного уравнения с частными производными.

**Ключевые слова:** группа Карно, пространство Соболева, теорема вложения, след функции, продолжение функций, теорема Уитни

### Похожие статьи:

- Н.Н. Романовский. Об оценках норм Бесова решений субэллиптических уравнений в трехмерном случае // Сиб. матем. журн., 52:5 (2011)
- С.К. Водопьянов, Н.А. Кудрявцева, "Нелинейная теория потенциала для пространств Соболева на группах Карно // Сиб. матем. журн., 50:5 (2009)
- И.М. Пупышев. Продолжение функций классов Соболева за границу области на группах Карно // Мат. тр., 10:2 (2007)
- Е.А. Плотникова. Интегральные представления и обобщенное неравенство Пуанкаре на группах Карно // Сиб. матем. журн., 49:2 (2008)

Рис. 2. Автоматически составляемая карточка публикации

Пользователь имеет возможности просмотреть незнакомые и заинтересовавшие его понятия. Эту функцию реализует сервис персонализированного составления карточек понятий.

Карточка понятия содержит: метаданные; различные наименования термина (на русском и английском языках); определение термина; ссылки на гипонимы/гиперонимы термина; ссылки на страницу термина на внешних ресурсах: OntoMath<sup>PRO</sup>, MSC 2010, Wolfram Mathworld, The Wolfram Functions Site, ScienceWISE, Математическая энциклопедия, Википедия и др.; список статей, релевантных заданному термину; сортировка статей осуществляется на основе известной меры TF-IDF и профиля пользователя.



**Определение:** Пространство функций, определенных на открытом множестве и интегрируемых с  $p$ -й степенью их модуля вместе со своими обобщенными производными до порядка  $m$  включительно.

**Внешние ресурсы:** OntoMath, ScienceWISE, MathWorld, Математическая энциклопедия, Википедия

#### Публикации:

- Л. Д. Кудрявцев, С. М. Никольский. Пространства дифференцируемых функций многих переменных и теоремы вложения // Анализ – 3, Итоги науки и техн. Сер. Соврем. пробл. мат. Фундам. направления, 26, ВИНИТИ, М., 1988
- А. А. Васильева. Достаточные условия вложения весового класса Соболева на области с условием Джона // Сиб. матем. журн., 56:1 (2015)
- С.К. Водопьянов, И.М. Пузышев. Следы функций из пространства Соболева на множествах Альфорса групп Карно // Сиб. матем. журн., 2007, том 48, номер 6
- Б.В. Трушин. Вложение пространства Соболева в пространство Орлича для области с нерегулярной границей // Матем. заметки, 2006, том 79, выпуск 5

Рис. 3. Автоматически составляемая карточка термина

**Семантический поиск по формулам** позволяет находить формулы, релевантные заданному математическому понятию. Релевантной считается формула, которая содержит переменную, обозначающую заданное понятие. При этом символическое представление переменной значения не имеет, что отличает семантический поиск от традиционного синтаксического (который отыскивает формулы по простому текстовому совпадению). Формулы извлекаются из текстов статей, ранжирование формул производится на основе TF-IDF, положение формулы в логической структуре документа (формула из формулировки теоремы или определения ранжируется выше, чем из доказательства), а также на основе профиля пользователя. Кроме того, поиск позволяет находить формулы, связывающие несколько понятий (например, давление и массу).

**Заключение.** Предложены архитектура и функции рекомендательного сервиса для профессиональных математиков. Сервис создается для автоматизации основных трудоемких процессов при работе с контентом. Главным отличием предложенной системы от существующих подходов к организации пользовательского интерфейса электронных коллекций являются предметно-ориентированные функции, такие, как персонализированные рекомендации статей и терминологии, реализуемые на основе семантических технологий. Ближайшие планы развития терминологического ресурса разрабатывающейся системы – включение понятий из [13] и установление семантических связей между ними.

Работа выполнена при финансовой поддержке РФФИ (проекты №№15-07-08522, 15-47-02472).

## **Литература**

1. *Bush V.* As we may think // The Atlantic Monthly. – Atlantic Media Co., Washington, 1945. – P. 101-108.
2. *Елизаров А. М., Липачев Е. К., Малахальцев М. А.* Веб-технологии для математика: Основы MathML. Практическое руководство. – М.: Физматлит, 2010. – 216 с.
3. *Елизаров А.М., Липачёв Е.К., Хохлов Ю.Е.* Семантические методы структурирования математического контента, обеспечивающие расширенную поисковую функциональность// Информационное общество. – 2013. – № 1-2. – С. 83-92.
4. *Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A., Sоловьев V.D., Zhiltsov N.G.* Mathematical knowledge representation: semantic models and formalisms // Lobachevskii J. of Mathematics. – 2014. – V. 35, No 4. – P. 347-353.
5. *Биряльцев Е.В., Елизаров А.М., Жильцов Н.Г., Липачёв Е.К., Невзорова О.А., Соловьев В.Д.* Методы анализа семантических данных математических электронных коллекций // Научно-техническая информация. Сер. 2. Информ. процессы и системы. – 2014. – №4. – С. 12-17.
6. *Ding L., Kolari P., Ding Z., Avancha S..* Using ontologies in the Semantic Web: a survey//Ontologies, Springer US. – 2007. – P. 79-113.
7. *Nilesh D., Kumar K., Pang B., Ramakrishnan R., Tomkins A., Bohannon P., Keerthi S., Merugu. S.* A Web of concepts// Proc. of the twenty-eighth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, ACM, 2009. – P. 1-12.
8. *Aberer K., Boyarsky A., Cudr-Mauroux P., Demartini G., Ruchayskiy O.* ScienceWISE: A Web-based interactive semantic platform for scientific collaboration // 10th Int. Semantic Web Conference (ISWC 2011 – Demo), 2011.
9. Textocat: катализатор текстовой аналитики. – URL: <http://textocat.com>
10. *Nevzorova O., Zhiltsov N., Zaikin D., Zhibrik O., Kirillovich A., Nevzorov V., Birialtsev E.* Bringing Math to LOD: a semantic publishing platform prototype for scientific collections in mathematics//12th Int. Semantic Web Conference, Sydney, NSW, Australia, October 21–25, 2013. Proceedings, Part I. 8218. Springer Berlin Heidelberg, 2013. – P. 379-394.
11. *Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E.* OntoMath PRO ontology: a linked data hub for mathematics // In Knowledge Engineering and the Semantic Web. Springer International Publishing. – 2014. – P. 105-119.
12. *Auer S., Bizer C., Kobilarov G., Lehmann J., Cyganiak R., Ives Z., Dbpedia: a nucleus for a Web of open data // In the Semantic Web.* – Springer Berlin Heidelberg, 2007. – P. 722-735.
13. *Математическая энциклопедия.* – М.: Советская энциклопедия. – Т. 1. – 1977; Т. 2. – 1979; Т. 3. – 1982; Т. 4. – 1984; Т. 5. – 1985.