

СЕМАНТИЧЕСКОЕ АННОТИРОВАНИЕ В СИСТЕМЕ УПРАВЛЕНИЯ ФИЗИКО-МАТЕМАТИЧЕСКИМ КОНТЕНТОМ

А. М. Елизаров¹, Н. Г. Жильцов², А. В. Кириллович², Е. К. Липачёв¹

*1 Институт математики и механики им. Н. И. Лобачевского
Казанского (Приволжского) федерального университета*

*2 Высшая школа информационных технологий и информационных систем
Казанского (Приволжского) федерального университета*

Введение. В настоящее время для научного сообщества одной из важных проблем, связанных с информатизацией исследовательской деятельности, является организация индивидуального информационного цифрового пространства ученого [1], а наиболее эффективный подход к решению проблем управления научной информацией обеспечивают семантические технологии, причем в результате автоматизации процессов обработки семантических связей формируется личное информационное пространство ученого.

Развитие методов автоматической обработки текстов позволило решать задачи извлечения знаний в терминах онтологий [2]. Онтологические модели предметных областей используются как технологическая основа построения рекомендательного сервиса, позволяющего выполнить персонализированный отбор научных документов в соответствии с семантическим профилем учёного [3].

Семантическое аннотирование. В настоящее время на платформе Science.tatarstan.ru (<http://science.tatarstan.ru/>) создается система аннотирования, основанная на онтологиях предметных областей. В исходный текст научной работы система автоматически добавляет ссылки на определения терминов – результатом является документ, содержащий помимо исходного текста научной работы блоки аннотаций, доступные через интерфейс пользователя. Отметим в связи с этим платформу ScienceWISE (<http://sciencewise.info>), с помощью которой пользователи ежедневно получают научную информацию, упорядоченную в соответствии с их личными интересами. При этом научные статьи ранжируются в соответствии с интересами пользователей, а на основе онтологий создается система закладок и аннотаций [4].

Подход, развиваемый нами, основан на использовании онтологий естественно-научных областей, прежде всего, онтологии математических знаний OntoMath^{PRO} [2, 5]. При формировании этой онтологии использовались методы онтологического моделирования с применением алгоритмов автоматического выделения понятий и связей [6]. Существенное значение для внедрения подобной системы имеет полнота «покрытия» онтологией терминологического массива электронной коллекции. В настоящее время в полуавтоматическом режиме проводится пополнение онтологии на основе терминологической базы Математической энциклопедии [7 – 9] и аннотаций, представленных на портале MathNet.ru [10].

Архитектура системы сформирована из нескольких модулей и представлена на рис.

1.

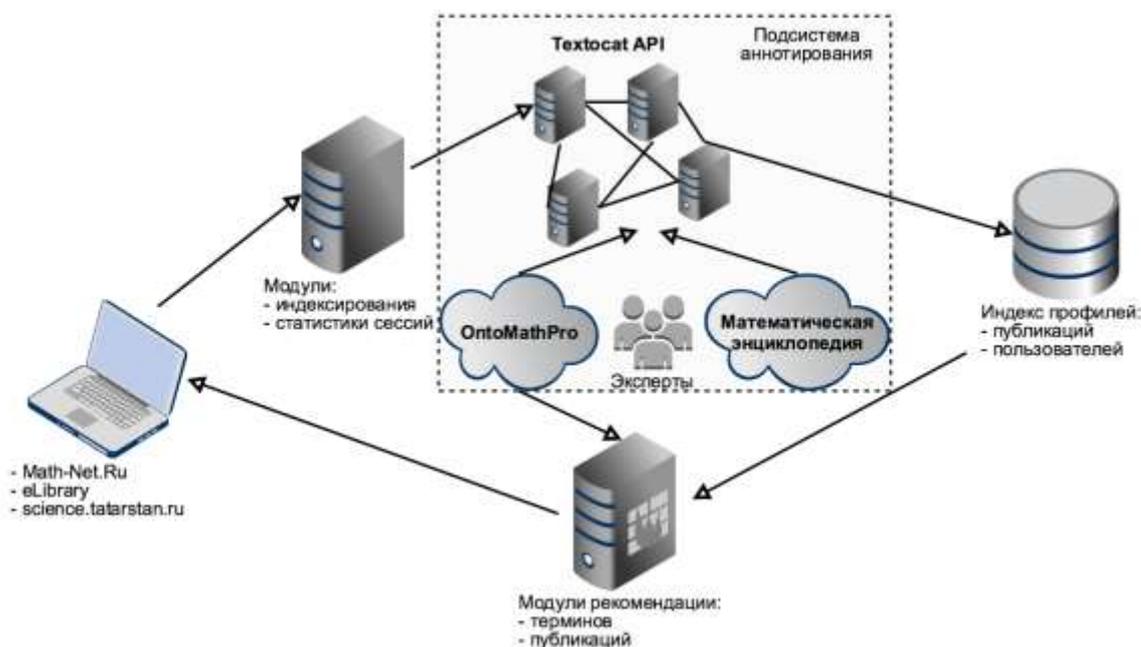


Рис. 1. Архитектура системы

Модуль семантического аннотирования производит связывание фраз из текста научных публикаций с концептами из терминологических источников, в качестве которых используются онтологии OntoMath^{PRO} и ScienceWISE, а также база знаний DBpedia. Сначала с помощью информации, полученной на основе лексико-синтаксического анализа, реализованного в облачном сервисе текстовой аналитики Textocat (<http://textocat.com>), генерируется набор фраз, рассматриваемых как кандидаты на связывание. При проведении такого анализа используются 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 следующем этапе модуль ранжирует концепты – кандидаты на связывание – с полученными фразами. При ранжировании учитываются характеристики, выражающие строковую близость имен кандидатов, а также контекст слов и других выделенных концептов, окружающих фразу-кандидата.

Далее происходит валидация высоко ранжированных кандидатов – каждому концепт-кандидату приписывается свое ранжирующее значение, рассматриваемое как «уверенность». В случае низкой уверенности модуля в кандидате на связывание для данной фразы связывание не производится.

Модуль составления семантического профиля пользователя формирует семантический профиль на основе онтологических концептов, исходя из истории просмотра пользователем статей на сформированном ресурсе: профиль содержит основные концепты, с которыми пользователю приходится работать наиболее часто, и отражает его компетенции.

Модуль рекомендации определений решает задачу составления рекомендаций при просмотре пользователем научных публикаций. Результатом являются список терминов, которые система, на основании сопоставления семантического профиля статьи и семантического профиля пользователя, выделила как наименее понятные пользователю.

Пользовательский интерфейс базируется на карточках публикаций (рис. 2), сформированных на портале MathNet.ru. Карточка публикации содержит название статьи, авторов, метаданные журнала, аннотацию статьи и, иногда, указанный автором список ключевых слов. Разработан специальный скрипт, результатом действия которого является новый раздел Extracted keywords, содержащий автоматически извлеченные ключевые слова со ссылками на

карточки терминов (рис. 3 и 4).

The screenshot shows a webpage header for 'Math-Net.Ru Sibirskii Matematicheskii Zhurnal'. Below the header is a navigation menu with links: JOURNALS, PEOPLE, ORGANISATIONS, CONFERENCES, SEMINARS, VIDEO LIBRARY, PERSONAL OFFICE. The main content area displays the following information:

Sibirsk. Mat. Zh., 2007, Volume 48, Number 6, Pages 1201–1221 (Mi smj1801)

This article is cited in 7 scientific papers (total in 7 papers)

Traces of Sobolev functions on the Ahlfors sets of Carnot groups

S. K. Vodop'yanov, I. M. Pupyshev

Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences

Abstract: We prove the converse of the trace theorem for the functions of the Sobolev spaces W_p^1 on a Carnot group on the real Ahlfors d -sets (the direct trace theorem was obtained in one of our previous publications). The theorem generalizes Johnsson's theorem on Sobolev functions on the Euclidean space. As a consequence we give a theorem on the boundary values of Sobolev functions on the boundary in a two-step Carnot group. We consider an example of application of the theorems to solvability of the boundary value problem for a partial differential equation.

Keywords: Carnot group, Sobolev space, embedding theorem, trace of a function, extension of functions, Whitney's theorem

● **Full text** (in Russian): PDF file (402 kB)
● **References** (in Russian): PDF file HTML файл

Рис. 2. Одна из карточек публикации на портале MathNet.ru

Отметим, что список ключевых слов имеет ряд ограничений. Во-первых, одно и то же понятие в разных статьях может обозначаться разными ключевыми словами. Во-вторых, ключевые слова не являются активными ссылками и не ведут на карточку соответствующего математического понятия. Наконец, во многих статьях список ключевых слов является неполным или даже вовсе отсутствует.

This screenshot is identical to the one in Figure 2, but it highlights the 'Extracted keywords' section. The text in this section is:

Extracted keywords: [Sobolev space](#), [Euclidean space](#), [Boundary value](#), [Partial differential equation](#), [Smooth boundary](#)

The other parts of the card, including the header, navigation menu, title, authors, abstract, and main keywords, are the same as in Figure 2.

Рис. 3. Результат работы сервиса – сформированный раздел Extracted keywords

Extracted keywords: [Sobolev space](#), [Euclidean space](#), [Boundary value](#),
[Partial differential equation](#), [Smooth boundary](#)

Рис. 4. Раздел Extracted keywords – автоматически извлеченные ключевые слова со ссылками на карточки терминов

Карточка термина представлена на рис. 5. Она содержит метаданные, различные наименования термина (на русском и английском языках), определение термина, ссылки на гипонимы/гиперонимы термина, ссылки на страницу термина на внешних ресурсах, а также ссылки на статьи с портала MathNet.ru, в которых использован данный термин.

Программная надстройка автоматически извлекает ключевые слова и добавляет блок автоматически извлеченных ключевых слов на карточку публикации. Каждое ключевое слово представляет собой активную ссылку на карточку соответствующего понятия. Ссылка ведет на одно и то же понятие, вне зависимости от того, какой конкретно термин использовался для обозначения этого понятия, и на каком языке.

Если авторы не указали список ключевых слов, то блок автоматически извлеченных ключевых слов является единственной заменой этого списка. Если указали, то его дополнением.

Нормированное пространство

Пространство Соболева

Пространство Соболева дробного порядка, Пространство Соболева бесконечного порядка

Определение: Пространство функций, определенных на открытом множестве Ω 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. Современ. пробл. мат. фундам. направления, 26, ВИНТИ, 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. Пупышев. 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. 5. Карточка одного из терминов

Работа выполнена при финансовой поддержке РФФИ (проекты 15-07-08522, 15-47-02472).

ЛИТЕРАТУРА

1. Моисеев Е.И., Муромский А.А., Тучкова Н.П., Меденников А.М. Организация персонального информационного пространства на основе модели предметной области автора // Науч-

- ный сервис в сети Интернет, многообразии суперкомпьютерных миров: труды Международной суперкомпьютерной конференции (г. Новороссийск, 22-27 сентября 2014 г.). – М.: Изд-во МГУ, 2014. – С. 429-431. – URL: <http://agora.guru.ru/abrau2014/pdf/429.pdf>.
2. Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A., Solovyev V.D., Zhiltsov N.G. Mathematical knowledge representation: semantic models and formalisms // *Lobachevskii J. of Mathematics*. – 2014. – V. 35, No 4. – P. 347-353.
 3. Елизаров А.М., Жильцов Н.Г., Иванов В.В., Кириллович А.В., Липачёв Е.К., Невзорова О.А. Семантический рекомендательный сервис в профессиональной деятельности математика // *Ученые записки Института социальных и гуманитарных знаний*. – 2015. – Вып. 1 (13). – С. 190-197.
 4. Astafiev A., Prokofyev R., Guéret C., Boyarsky A., Ruchayskiy O. ScienceWISE: A Web-based interactive semantic platform for paper annotation and ontology editing // In: *Extended Semantic Web Conference 2012, Greece, 2012*. URL: http://2012.eswc-conferences.org/sites/default/files/eswc2012_submission_316.pdf.
 5. Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E. OntoMathPRO ontology: a linked data hub for mathematics // In *Knowledge Engineering and the Semantic Web*. Springer International Publishing. – 2014. – P. 105-119.
 6. 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.
 7. Математическая энциклопедия. – М.: Советская энциклопедия. – Т. 1. – 1977; Т. 2. – 1979; Т. 3. – 1982; Т. 4. – 1984; Т. 5. – 1985.
 8. Математическая энциклопедия. URL: http://gufo.me/matenc_a
 9. Encyclopedia of Mathematics. URL: http://www.encyclopediaofmath.org/index.php/Main_Page
 10. Общероссийский математический портал. URL: <http://www.mathnet.ru/>