



КРАТКИЕ СООБЩЕНИЕ

УДК 004.6

Кластеризация и многоступенчатый семантический анализ научных текстов

*Исаков С.С.**

Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация

ORCID: <https://orcid.org/0000-0003-1719-2355>

e-mail: isakovss@mgppu.ru

На сегодняшний день многие проблемы компетентного подхода остаются не решенными как в методологическом и методическом плане, так и в плане организации деятельности и создания соответствующих технологий. Для формирования научно-исследовательских компетенций университета соответствующие сведения о проектах, научных работах и специализации сотрудников университета необходимо собирать и обрабатывать, что в современных условиях при большом разнообразии научных направлений и количестве отдельных работ и проектов требует создания специализированной информационной системы. Задача выявления компетенций сотрудников научной организации может быть решена путем кластеризации текстов их научных работ.

Ключевые слова: информационная система, формирование компетенций, научно-исследовательская компетенция, ключевые слова, научная специальность, научно-исследовательский проект, специализация сотрудников, веб-ресурс, реляционная база данных, формальная семантика.

Для цитаты:

Исаков С.С. Кластеризация и многоступенчатый анализ научных текстов // Моделирование и анализ данных. 2022. Том 12. № 4. С. 105–109. DOI: <https://doi.org/10.17759/mda.2022120408>

**Исаков Сергей Сергеевич*, преподаватель, аспирант, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0003-1719-2355>, e-mail: isakovss@mgppu.ru



1. ПОСТАНОВКА ЗАДАЧИ

Известно, что перед высшей профессиональной школой сегодня поставлена задача подготовки выпускников к проведению научных исследований, что в терминах компетентностного подхода звучит как формирование научно-исследовательской компетентности.

На сегодняшний день многие проблемы компетентностного подхода остаются не решенными как в методологическом и методическом плане, так и в плане организации деятельности и создания соответствующих технологий. Научно-исследовательские компетенции университета в первую очередь определяются составом его сотрудников, преподавателями и научными работниками, которые могут осуществлять различные научно-исследовательские работы на должном уровне. Поэтому научно-исследовательские компетенции университета, по-видимому, следует и можно определять по тем проектам и научно-исследовательским работам, которые осуществляет университет с учетом общей квалификации его научного персонала (образования и специализации сотрудников). При этом профессиональная направленность этих компетенций так или иначе связана с известными научными специальностями научных сотрудников. Для формирования научно-исследовательских компетенций университета соответствующие сведения о проектах, научных работах и специализации сотрудников университета необходимо собирать и обрабатывать, что в современных условиях при большом разнообразии научных направлений и количестве отдельных работ и проектов требует создания специализированной информационной системы.

Цель работы по созданию веб-ресурса для формирования научно-исследовательских компетенций заключалась в разработке компьютерной информационной системы для определения научно-исследовательских компетенций МГППУ и планирования научной деятельности высшего учебного заведения с функциями частичной автоматизации формирования описаний научно-исследовательских компетенций.

Задачи проекта включали: анализ вопросов формализации описаний научно-исследовательских компетенций; определение функциональных требований к ИС и составление технического задания на разработку ИС; синтаксический и семантический анализ структуры и алгоритмизация отношений между составляющими научно-исследовательских компетенций; разработка механизмов реляционной базы данных научной специализации и научно-исследовательских работ и проектов университета; синтаксический и семантический анализ структуры и алгоритмизация отношений между составляющими научно-исследовательских компетенций; разработка механизмов реляционной базы данных научной специализации и научно-исследовательских работ и проектов университета; разработка алгоритмов автоматизации формирования описаний научно-исследовательских компетенций.

Такая информационная система может обеспечить систематизацию, упорядочение и автоматизацию сведений о научно-исследовательских компетенциях университета и упрощение работы с этой информацией, даст возможность лучше



координировать и планировать дальнейшую научно-исследовательскую деятельность университета и осуществлять экспертизу постановки новых научных задач, позволит провести рекламу и продвижение проектов университета. Это в результате может дать университету необходимые конкурентные преимущества на рынке. В настоящее время стандартных информационных систем такого типа на рынке не существует, что делает проект по разработке информационной системы «научно-исследовательские компетенции университета» актуальным.

2. ТЕХНИЧЕСКАЯ РЕАЛИЗАЦИЯ

Программная реализация информационной системы представляет из себя веб-ресурс, хранящийся на сервере МГППУ. В качестве приложения-сервера используется широко распространенное приложение Apache. Система написана с использованием языков: PHP, JavaScript, CSS, HTML. В качестве приложения базы данных используется приложение MySQL. Так как разрабатываемый веб-ресурс является информационной системой сложной структуры, то для разработки используется фреймворк Laravel, включающий в себя фреймворк Symfony.

В качестве языка представления шаблонов графических HTML элементов в системе используется язык Twig. Гибкое и эффективное взаимодействие с базой данных совершается с использованием библиотеки Eloquent. Вся передача данных внутри системы и от сервера к клиентскому приложению производится на языке структуризации данных JSON. В целом архитектура приложения создана на шаблоне проектирования MVC (Model View Controller).

Шаблон проектирования MVC предполагает разделение данных приложения, пользовательского интерфейса и управляющей логики на три отдельных компонента: Модель, Представление и Контроллер – таким образом, что модификация каждого компонента может осуществляться независимо. Такой систематизированный подход позволяет в достаточной степени и с наилучшей эффективностью использовать концепцию объектно-ориентированного программирования, основанную на представлении программы в виде совокупности объектов, каждый из которых является экземпляром определённого класса, а классы образуют иерархию наследования.

3. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Задача выявления компетенций сотрудников научной организации может быть решена путем кластеризации текстов их научных работ. При этом необходимо учитывать наличие у сотрудников утвержденной научной квалификации и их распределение по научным специальностям соответствующей номенклатуры. Наиболее простой вариант решения задачи сводится к формализованному представлению текстов с последующей оценкой их семантической близости. Связь представлений текстов и научных специальностей представляется матрицей в метрике TF-IDF.

Предлагаемый подход предполагает вычисление оценок согласованности документов с ключевыми фрагментами описаний научных специальностей. Размерность пространства, в котором представлены эти тексты, понижается с помощью сингулярного разложения матриц в метрике TF-IDF. Это обеспечивает понижение размерности задачи и удаление несущественной информации. Последующее многомерное шкалирование позволяет получить представления научных текстов в пространстве малой размерности, что является исходными данными для последующей кластеризации. Сингулярное разложение матрицы, связывающей сотрудников и кластеры научных текстов, с последующим многомерным шкалированием позволяет выполнить кластеризацию носителей компетенций.

Литература

1. *Андреев А.В., Митрофанова О.А., Соколов К.В.* Введение в формальную семантику. СПб.: СПбГУ. РИО. Филол. фак., 2014. 88 с.
2. *Бах Э.* Неформальные лекции по формальной семантике. М.: УРСС, 2010. 224 с.
3. *Гуружапов В.А., Марголис А.А.* Проектирование модели практико-ориентированной подготовки педагогических кадров по программам бакалавриата по направлению подготовки «Психолого- педагогическое образование (Учитель начальных классов)» на основе сетевого взаимодействия образовательных организаций, реализующих программы высшего образования и начального общего образования // Психологическая наука и образование. 2014. Т.19. № 3. С.143–159.
4. *Хамел, Г.* Конкурируя за будущее. Создание рынков завтрашнего дня / Г. Хамел, К.К. Прахалад; пер. с англ. М.: ЗАО «Олимп-Бизнес», 2002. 288 с.
5. *Хван А.А.* Теоретические и прикладные проблемы измерения компетенций // В сборнике: Оценка качества обучения в образовательных учреждениях. Уральский государственный педагогический университет. 2012. С. 105–110.
6. *Хисамиева Л.Г.* Компетенции и компетентность в структуре научно-исследовательской деятельности // Вестник ОГУ. 2014. № 9 (170). С. 33–37.



Clustering and Multistage Semantic Analysis of Scientific Texts

Sergey S. Isakov *

Moscow State University of Psychology and Education (MSUPE), Moscow, Russia

ORCID: <https://orcid.org/0000-0003-1719-2355>

e-mail: isakovss@mgppu.ru

To date, many problems of the competency-based approach remain unresolved both in terms of methodology and methodological, and in terms of organizing activities and creating appropriate technologies. To form the research competencies of the university, relevant information about projects, scientific work and specialization of university staff must be collected and processed, which in modern conditions, with a wide variety of scientific areas and the number of individual works and projects, requires the creation of a specialized information system. The task of identifying the competencies of employees of a scientific organization can be solved by clustering the texts of their scientific works.

Keywords: information system, competence formation, research competence, keywords, scientific specialty, research project, employee specialization, web resource, relational database, formal semantics.

For citation:

Isakov S.S. Clustering and Multistage Semantic Analysis of Scientific Texts. *Modelirovanie i analiz dannykh = Modelling and Data Analysis*, 2022. Vol. 12, no. 4, pp. 105–109. DOI: <https://doi.org/10.17759/mda.2022120408> (In Russ., abstr. in Engl.).

References

1. Andreev A.V., Mitrofanova O.A., Sokolov K.V. Introduction to formal semantics. St. Petersburg: St. Petersburg State University. RIO. Philol. faculty, 2014. 88 p.
2. Bach E. Informal lectures on formal semantics. M.: URSS, 2010. 224 p.
3. Guruzhapov V.A., Margolis A.A. Designing a model of practice-oriented training of teaching staff in undergraduate programs in the direction of training “Psychological and pedagogical education (Primary school teacher)” based on the network interaction of educational organizations implementing programs of higher education and primary general education. *Psychological Science and Education*. 2014. Vol.19. No. 3. P. 143–159.
4. Hamel, G. Competing for the future. Creating the markets of tomorrow. G. Hamel, K.K. Prahalad; per. from English. M.: CJSC “Olimp-Business”, 2002. 288 p.
5. Khvan A.A. Theoretical and applied problems of measuring competencies. In the collection: Evaluation of the quality of education in educational institutions. Ural State Pedagogical University. 2012. P. 105–110.
6. Khisamieva L.G. Competence and competence in the structure of research activities. *Bulletin of OSU*. 2014. No. 9 (170). pp. 33–37.

***Sergey S. Isakov**, Lecturer, Postgraduate student, Moscow State University of Psychology and Education (MSUPE), Moscow, Russia, ORCID: <https://orcid.org/0000-0003-1719-2355>, e-mail: isakovss@mgppu.ru

Получена 19.11.2022

Принята в печать 19.12.2022

Received 19.11.2022

Accepted 19.12.2022