
Секция 1 | ИНЖЕНЕРИЯ ЗНАНИЙ

УДК 004.8

doi: 10.15622/rcai.2025.003

КАРТОГРАФИЯ И СЕМАНТИКА НАУЧНОГО ЗНАНИЯ: ПИЛОТНЫЙ ПРОЕКТ

Т.А. Гаврилова (*gavrilova@gsom.spbu.ru*)^A

В. Шванкин (*v.shvankin@salesai.ru*)^B

М.В. Кубельский (*m.kubelskiy@gsom.spbu.ru*)^A

Н.В. Иваникова (*n.ivanikova@spbu.ru*)^C

В. Луцков (*st098065@student.spbu.ru*)^A

^A Высшая школа менеджмента,
Санкт-Петербургский государственный университет,
Санкт-Петербург

^B ООО “Герофарм”, Санкт-Петербург

^C Управление научных исследований,
Санкт-Петербургский государственный университет,
Санкт-Петербург

В докладе представлена методика визуализации библиометрических данных исследователей одного из подразделений университета на основе семантики текстов публикаций. С помощью глубокого рассмотрения метаданных публикаций, включающего семантический анализ названий и аннотаций, сформирован прототип интерактивных карт научных интересов и связей между авторами. Такой подход позволяет выявить ключевые направления исследований, междисциплинарные взаимодействия и структуру научного коллектива, что способствует более эффективному управлению научной деятельностью и развитию сотрудничества внутри университета.

Ключевые слова: карты знаний, библиометрия, семантический анализ.

Введение

Интеллектуальный капитал университета, как сумма всех нематериальных активов, включает в себя в первую очередь человеческий капитал, т.е. знания и навыки преподавателей, их публикации и выступления [Полушкевич, 2018]. Долгое время этот капитал трудно поддавался изменению и какой-либо автоматизации в управлении [García-Carbonell et al., 2021], [Гаврилова и др., 2010]. Анализ библиометрических данных является важным инструментом управления и мониторинга этого актива [Руководство по наукометрии: индикаторы развития науки и технологии, 2021] в наукоемких коллективах [Donthu et al., 2021].

Специальное программное обеспечение, представляющее результаты библиометрического анализа в графическом виде, делает его доступным для широкого круга пользователей (VOSviewer, CiteSpace, SciVal, Bibliometrix / Biblioshiny, eLIBRARY). Визуализация результатов анализа позволяет создавать комплексные карты научного знания с учетом семантики научных текстов. Автоматизированный семантический анализ определенных разделов научных публикаций, таких как, например, название и аннотация, может обеспечить объективное представление об интеллектуальных активах, которыми обладает научно-исследовательское или образовательное учреждение [Chagnon et al., 2024].

В докладе используются и развиваются некоторые из результатов проекта «Методология и технология разработки цифровых карт знаний для учебных и научных коллективов (МЕТАКАРТА)» (грант РФФИ 2023-24 № 23-21-00168).

1. Теоретические основы картографии научного знания

Библиометрический анализ научных публикаций является ценным инструментом для оценки результатов исследований, влияния и тенденций [Donthu et al., 2021], [McAlliste et al., 2022]. Для эффективной передачи сложных данных и взаимосвязей, выявленных с помощью библиометрии, визуальные диаграммы играют решающую роль. Они облегчают идентификацию закономерностей, упрощают коммуникацию и помогают в принятии решений [Дудко и др., 2023].

Идея визуализации сетей цитирования и картирования науки восходит к 60-м годам 20-го века и тесно связана с развитием наукометрии. Один из основоположников наукометрии Д. Прайс предложил на основе анализа сетей цитирования документов определять фронты развития научных исследований [Price, 1965]. В настоящее время при построении библиометрических карт чаще всего используют анализ цитирования и со-встречаемости ключевых слов в документах. Метод ко-цитирования основан на выявлении публикаций, которые совместно цитируются другими авторами [Маршак-ова, 1973], Small, 1973]. При кластеризации на основе совместного использо-

вания ключевых слов используют автоматическое извлечение ключевых слов из текстов или предоставленные авторами ключевые слова [Van Eck et al., 2014]. Современные подходы в области обработки естественного языка и кластеризации векторных представлений текстов, такие как тематическое моделирование – мощный инструмент, создающий новые возможности для картирования науки [Thijs et al. 2018].

Основными типами библиометрических визуальных диаграмм являются [Subramanyam, 1983], [Судакова и др., 2025]:

- Сети цитирования: визуализируют связи между публикациями по цитированию.
- Сети совместных ключевых слов: раскрывают темы и взаимосвязи.
- Диаграммы влияния цитирования: представляют количество цитирований и импакт-факторы публикаций, позволяя сравнивать высокоцитируемые работы.
- Карты сотрудничества: иллюстрирует отношения на основе соавторства.
- Карты предметных категорий на основе существующих рубрикаторов: иерархически организуют научные публикации и их взаимосвязи по категориям.

Многие авторы обсуждают важность полноты и точности данных, которая может повлиять на достоверность визуализаций [Руководство по наукометрии: индикаторы развития науки и технологии, 2021]. Следует также учитывать субъективность интерпретации и особенности публикационной активности в изучаемой области знания. Кроме того, многие исследователи сталкиваются с ограничениями программного обеспечения [Moral-Munoz et al, 2020].

2. Методология построения карт научного знания

Пилотная разработка библиометрических карт на основе семантического анализа потребовала решения ряда задач в рамках концепции (proof-of-concept), т.е. доказательства жизнеспособность выбранного подхода.

Для этого в рамках исследования были выделены следующие этапы (подробнее описаны в следующих параграфах):

- a) Выбор показателей и метрик.
- b) Анализ и отбор источников данных.
- c) Сбор и описание данных.
- d) Разработка прототипа архитектуры программного обеспечения.
- e) Анализ и визуализация данных.
- f) Создание сценариев использования.

а) Выбор показателей

В работе использовался реестр авторов в привязке к кафедрам. Он содержит ссылки на профили сотрудников на основных библиографических ресурсах. Реестр включает в себя справочную информацию о сотрудниках и содержит наборы полей: *ФИО/Факультет/Должность/Степень/Elibrary_SPIN/Scopus_AuthorID/WoS_ResearcherID/WoS_link*.

Базовыми показателями, используемыми для анализа результативности ученых, являются количество публикаций и цитирований, индекс Хирша.

При анализе библиометрических данных важно учитывать отличия в характере публикационной активности в различных областях знаний. Например, в общественных и гуманитарных науках влияние книжных, а также не англоязычных публикаций, многие из которых не индексируются в международных наукометрических базах данных, выше, чем в естественных науках [Руководство по наукометрии: индикаторы развития науки и технологии, 2021], [Aksnes et al., 2021]. Таким образом, для увеличения охвата публикаций и более полного анализа следует обращаться в том числе к данным из национальных указателей цитирования, включающих не англоязычные публикации [Sile et al., 2017]. При этом, наряду с публикациями в периодических научных изданиях, целесообразно включать и другие типы публикаций, в частности монографии и сборники материалов конференций.

Для обеспечения возможности построения различных карт знаний были сформулированы минимальные требования к данным. Они представлены в табл. 1.

Таблица 1

Метаданные и показатели

А. Мета-данные автора /участника	Б. Библиометрические показатели по каждому автору/сотруднику	С. Семантически значимая информация для определения областей компетенции
<ul style="list-style-type: none">• Фамилия• Имя• Уникальный идентификатор	<ul style="list-style-type: none">• Список публикаций• Количество цитирований для каждой публикации• Индекс Хирша и другие имеющиеся агрегированные показатели	<ul style="list-style-type: none">• Названия публикаций• Аннотации публикаций

б) Анализ источников данных

В качестве источников данных о результатах научной деятельности сотрудников были рассмотрены следующие ресурсы сети Интернет:

- РИНЦ (<https://elibrary.ru/>).

- Scopus (<https://www.scopus.com/home.uri>).
- Web of Science (<https://clarivate.com/>).
- Google Scholar (<https://scholar.google.com/>).

Для выбора источника данных были сформированы следующие критерии: (а) возможность автоматической загрузки данных / (б) свободный доступ / (в) полнота данных / (г) учет публикаций и на английском, и на русском языке.

Библиографические базы данных Scopus, Web of Science и РИНЦ не соответствуют критериям (а) и (б), так как для получения доступа к API системы требуется оплата подписки. В Scopus и Web of Science не англоязычные публикации недостаточно представлены [Aksnes et al., 2021], а значит не соблюдается еще и критерий (г).

Google Scholar, представляющий собой автоматизированный агрегатный сервис, осуществляющий сбор метаданных публикаций путем сканирования ресурсов сети Интернет, обеспечивает достаточно высокую полноту охвата публикаций. При этом данные Google Scholar находятся в свободном доступе с возможностью их автоматической загрузки.

Таким образом, в качестве источника данных был выбран Google Scholar, отвечающий большинству критериев.

с) Разработка прототипа архитектуры программного обеспечения

Для автоматического сбора данных, их анализа и визуализации карт знаний был разработан программный модуль BIB-METR.2 на языке Python, включающий библиотеки для поиска и подключения к прокси Google Scholar, парсинга и обработки данных [Kubelski et al., 2024], семантического анализа и создания графов – scholarly, pandas, PyVis, NetworkX, spaCy, Pickle, SentenceTransformers и другие. Разработка оригинального модуля позволила обеспечить гибкость работы с необходимой структурой данных и создала платформу для дальнейших этапов исследования.

д) Анализ и визуализация данных

BIB-METR.2 реализует систему семантического анализа и визуализации научной активности. Результатом работы модуля стала интерактивная карта знаний, основанная на библиометрических и текстовых данных, агрегированных по кафедрам, научным сотрудникам и тематическим направлениям.

Построенная карта знаний обладает четко выраженной структурной иерархией, обеспечивая возможность анализа научной активности на нескольких уровнях детализации. Визуализация выполнена с использованием графовой модели (рис. 1), где:

- кафедры представлены в виде прямоугольных узлов синего цвета;
- исследователи (авторы) – в виде жёлтых узлов круглой формы, размер которых зависит от количества публикаций и уровня цитируемости;
- ключевые исследовательские термины – в виде небольших красных узлов.

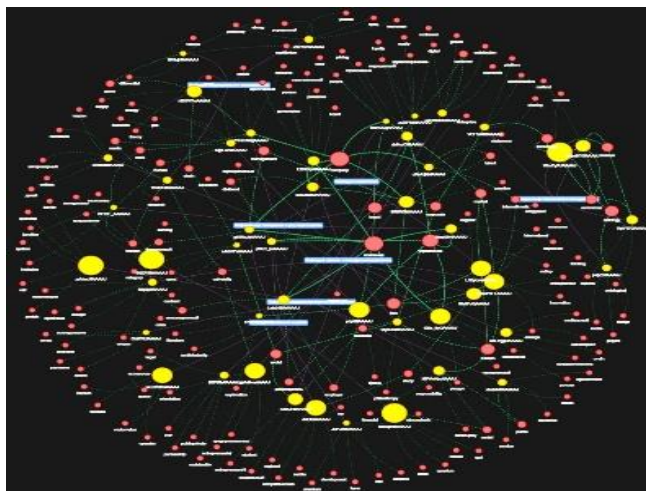


Рис. 1. Исследователи и их научные интересы

Таким образом, достигается высокая степень наглядности, что способствует быстрой идентификации центров научной активности, ключевых участников и направлений научной работы. Кроме того, визуализация позволяет выявлять междисциплинарные взаимодействия и исследовательские пробелы, ранее не поддававшиеся прямой аналитике.

Модуль обеспечивает интерактивную работу с данными, включая фильтрацию, детализацию связей, переход к исходной публикационной информации, что расширяет его аналитический потенциал и делает его применимым как для внутренних нужд вуза, так и для внешних заинтересованных сторон.

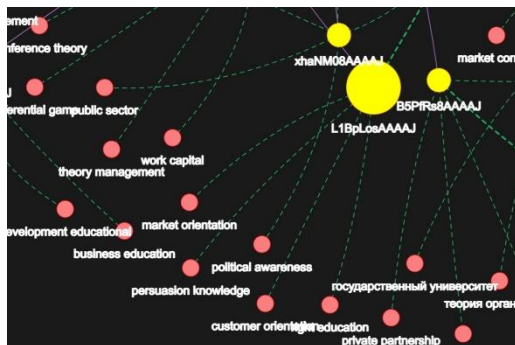


Рис. 2. Увеличенный фрагмент карты знаний

е) О сценариях использования

Ниже представлены сценарии для различных групп пользователей:

е1) Для исследователей система служит инструментом анализа текущих научных трендов и поиска перспективных направлений. Карта позволяет:

- выявить коллег, работающих в смежных тематических областях;
- формировать междисциплинарные научные коллективы;
- оценить собственную публикационную активность и позиционирование в научном сообществе университета.

е2) Руководство вуза получает возможность принимать обоснованные стратегические решения на основе агрегированной информации о:

- ведущих научных направлениях;
 - ключевых исследовательских группах;
 - динамике публикационной активности и цитируемости.
- Формализованные аналитические запросы, доступные в системе:

- Определение авторов, объединяющих несколько кафедр.
- Идентификация направлений с единственным специалистом.
- Выявление направлений с высокой динамикой публикационной активности.

е3) Для сторонних пользователей (бизнеса, государственных структур, фондов) карта знаний выступает в роли инструмента:

- быстрой оценки научного потенциала университета;
- выбора исполнителей под прикладные научные задачи;
- принятия решений о финансировании и развитии партнёрств.

Таким образом, реализация модуля BIB-METR.2 создаёт предпосылки для формирования единой аналитико-коммуникационной платформы, способствующей эффективному взаимодействию между всеми заинтересованными субъектами: научным сообществом, администрацией вуза и внешними партнёрами.

Заключение

В докладе обсуждаются принципы разработки и предварительные результаты пилотного проекта "НАучная ВИзуализация для сотрудничества и развития" (НАВИЯ). Проект ориентирован на создание методологии разработки библиометрических карт знаний кафедр и факультетов университета. В ходе работы над проектом предложен и апробирован подход к обобщению аналитической информации о публикационной активности научно-педагогических работников вуза. Дальнейшие направления развития предлагаемого подхода включают использование прототипа для расширения палитры анализа и углубления семантического анализа для выявления неочевидных связей и ассоциаций.

Основные бенефициары проекта и решаемые задачи:

- Сами члены научного коллектива и отдельные исследователи: выявление пробелов и ниш в исследованиях, поиск партнеров для сотрудничества, оценка эффективности, мониторинг трендов.
- Руководство университетов и научных институтов: принятие решений о финансировании, развитии приоритетных направлений, оценке результативности научных групп.
- Государственные органы и бизнес: определение приоритетных направлений развития науки и технологий, оценка вклада отдельных ученых и организаций в развитие страны.

Список литературы

- [Гаврилова и др., 2010] Гаврилова Т.А., Кудрявцев Д.В. Информационные технологии управления знаниями // В книге: Инновационное развитие: экономика, интеллектуальные ресурсы, управление знаниями. – М.: ИНФРА-М, 2009. – С. 500-515.
- [Дудко и др., 2023] Дудко В.В., Патаракин, Е.Д. Исследование научных школ университета средствами библиометрического картирования. Территория новых возможностей // Вестник Владивостокского государственного университета экономики и сервиса. – 2023. – Т. 15, № 1. – С. 150-167. – doi: 10.24866/VVSU/2949-1258/2023-1/150-167.
- [Маршакова, 1973] Маршакова И.В. Система связей между документами, построенная на основе ссылок: по данным Science Citation Index // Научно-техническая информация. Сер. 2. – 1973. – № 6. – С. 3-8.
- [Полушкевич, 2018] Полушкевич О.А. Интеллектуальный капитал университета // Alma mater (Вестник высшей школы). – 2018. – № 5. – С. 25-27. – doi: 10.20339/AM.05-18.025.
- [Руководство по наукометрии: индикаторы развития науки и технологий, 2021] Акоев М.А., Маркусова В.А., Москалева О.В., Писляков В.В. Руководство по наукометрии: индикаторы развития науки и техники. – М.: Изд-во Уральского университета, 2021. – 358 с. – doi: 10.15826/B978-5-7996-1352-5.0000.
- [Судакова и др., 2025] Судакова А.Е., Агарков Г.А. Датасет о наукометрии российских ученых: кейс е. Library // Вопросы образования. – 2025. – № 1. – С. 304-330. – doi: 10.17323/vo-2025-21514.
- [Aksnes et al., 2021] Aksnes D.W., Sivertsen G. A Criteria-based Assessment of the Coverage of Scopus and Web of Science // Journal of Data and Information Science. – 2019. – Vol. 4(1). – P. 1-21. – doi: 10.2478/jdis-2019-0001.
- [Donthu et al., 2021] Donthu N., Kumar S., Mukherjee D., Pandey N., and Lim W.M. How to conduct a bibliometric analysis: An overview and guidelines // Journal of business research. – 2021. – No. 133. – P. 285-296. – doi: 10.1016/j.jbusres.2021.04.070.
- [García-Carbonell et al., 2021] García-Carbonell N., Guerrero-Alba F., Martín-Alcázar F., Sánchez-Gardey G. Academic human capital in universities: definition and proposal of a measurement scale. Science and public policy. – 2021. – Vol. 48(6), 877. – doi: 888.10.1093/scipol/scab062.

- [**Kubelskiy et al., 2024**] Kubelskiy M., Kuznetsova A., Leshcheva I., Gavrilova T., Shvankin V. Ontology-Based Approach for Research Activity Mapping // International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Novosibirsk, Russian Federation. – 2024. – P. 334-338. – doi: 10.1109/SIBIRCON63777.2024.10758489.
- [**McAllister et al., 2022**] McAllister J.T., Lennertz L., Mojica Z.A. Mapping a discipline: a guide to using VOSviewer for bibliometric and visual analysis. *Science & Technology Libraries*. – 2022. – Vol. 41, No. 3. – P. 319-348. – doi: 10.26811/peuradeun.v12i3.1125.
- [**Moral-Munoz et al., 2020**] Moral-Muñoz J.A., Herrera-Viedma E., Santisteban-Espejo A., and Cobo M.J. Software tools for conducting bibliometric analysis in science: An up-to-date review // *Information Professional*. – 2020. – 29(1). – doi: 10.3145/epi.2020.ene.03.
- [**Van Eck et al., 2014**] Van Eck N.J., Waltman L. Visualizing bibliometric networks. In *Measuring scholarly impact: Methods and practice*. – Springer, 2015. – P. 285-320. – doi: 10.1007/978-3-319-10377-8_13.
- [**Price, 1965**] Price D.J.D.S. Networks of Scientific Papers // *Science*. – 1965. – Vol. 149(3683). – P. 510-515. – doi: 10.1126/science.149.3683.510.
- [**Chagnon et al., 2025**] Chagnon E., Pandolfi R., Donatelli J., Ushizima D. Benchmarking topic models on scientific articles using BERTeity // *Natural Language Processing Journal*. – 2024. – Vol. 6(100044). – doi: 10.1016/j.nlp.2023.100044.
- [**Sile et al., 2017**] Sile L., Guns R., Sivertsen G., Engels T. European Databases and Repositories for Social Sciences and Humanities Research Output. Report. Antwerp: ECOOM & ENRESSH. – 2017. – doi: 10.6084/m9.figshare.5172322 2017.
- [**Small, 1973**] Small H. Co-Citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents // *Journal of the American Society for Information Science*. – 1973. – doi: 10.1002/asi.4630240406.
- [**Subramanyam, 1973**] Subramanyam K. Bibliometric studies of research collaboration: A review // *Journal of information Science*. – 1983. – 6(1). – P. 33-38. – doi: 10.1177/016555158300600.
- [**Thijs et al., 2018**] Thijs B., Glanzel W. The contribution of the lexical component in hybrid clustering, the case of four decades of “Scientometrics” // *Scientometrics*. – 2018. – Vol. 115. – P. 21-33. – doi: 10.1007/s11192-018-2659-0.