



DOI: 10.19181/4m.2025.34.1.2

EDN: RBNLBE

**ТЕХНОЛОГИЯ АНАЛИЗА БОЛЬШИХ ДАННЫХ
ДЛЯ ВЫЯВЛЕНИЯ ОРГАНИЗАЦИОННЫХ СВЯЗЕЙ:
ВОЗМОЖНОСТИ СИСТЕМЫ iFORA**

Ирина Владимировна Логинова

Национальный исследовательский университет «Высшая школа
экономики», Москва, Россия

Эл. почта: iloginova@hse.ru

ORCID: 0000-0002-3376-2728

Анна Сергеевна Пиекалнитс

Национальный исследовательский университет «Высшая школа
экономики», Москва, Россия

Эл. почта: apiekalnits@hse.ru

ORCID: 0000-0003-0585-5350

Фёдор Михайлович Грозовский

Национальный исследовательский университет «Высшая школа
экономики», Москва, Россия

Эл. почта: fgrozovskiy@hse.ru

ORCID: 0009-0000-4342-7942

Для цитирования: *Логинова И. В., Пиекалнитс А. С., Грозовский Ф. М.*
Технология анализа больших данных для выявления организационных связей:
возможности системы iFORA // Социология: методология, методы, мате-
матическое моделирование (Социология: 4М). 2025. № 60. С. 36-88.
DOI: 10.19181/4m.2025.34.1.2. EDN: RBNLBE.

Цель настоящей статьи заключается в демонстрации подходов по анализу связей центров компетенций с использованием системы интеллектуального анализа больших текстово-документных данных iFORA на примере картирования области «Промышленные информационно-управляющие и телекоммуникационные системы». Результаты исследования включают апробацию реализованных в системе iFORA подходов и методов к задачам, которые традиционно решаются в рамках организационного сетевого анализа. Основным результатом проведенного анализа – граф связей потенциальных центров компетенций в области «Промышленные информационно-управляющие и телекоммуникационные системы», выявленных из научных публикаций с использованием инструментов семантического анализа, с отражением связей между ними и их интерпретацией – формированием перечней гипотез по существующим кооперациям, фиксируемым на основе этих связей. Приводятся предложения по направлениям использования исследования связей центров компетенций на основе данных семантического анализа больших текстовых массивов информации.

Ключевые слова: анализ данных, большие данные, семантический анализ, интеллектуальный анализ данных, текст-майнинг, организационный сетевой анализ, выявление коопераций.

Благодарности: статья подготовлена в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики».

Постановка исследовательской задачи

Сетевой анализ представляет собой оценку взаимоотношений между анализируемыми единицами, которая позволяет выявить положение объекта в общей сети, а также определить структуру взаимодействия с другими объектами [1]. Данное направление анализа зародилось в 1930-х гг. с выходом в печать книги Я. Морено «Кто останется в живых?» [2], в которой был применен метод сетевого анализа для описания взаимоотноше-

ний среди заключенных. Метод получил широкое применение в последние годы, что связано с активным развитием как методологических и технических подходов, так и доступных программных решений для обработки данных в рамках сетевого анализа [3].

Одной из наиболее динамично развивающихся сфер исследований в области сетевого анализа является организационный сетевой анализ (Organizational Network Analysis — ONA; далее — ОСА) [8]. В широком смысле ОСА можно определить в качестве одного из методов организационной диагностики, позволяющей провести систематический сбор и анализ информации об организации для выявления проблем функционирования и способов их преодоления [13; 14]. В более узком смысле данный метод анализа представляет собой набор аналитических инструментов для отображения связей между работниками, группами работников, отделами или организациями в ходе выполнения работы. Организационный сетевой анализ может применяться для решения задач социологических исследований, в т. ч. как инструмент мониторинга слабых сигналов и раннего выявления потенциальных корпоративных и социальных связей между людьми, аффилированными с организациями.

На сегодняшний день ОСА является востребованным методом выявления взаимосвязей внутри организации (например, на уровне отдельных офисов, департаментов, дочерних организаций), который позволяет увеличить операционную эффективность, а также повысить эффективность управленческих бизнес-процессов. Современные программные решения позволяют сделать процесс ОСА более эффективным за счет возможности анализа больших массивов данных и продвинутых способов визуализации.

Так, ОСА в значительной степени опирается на использование графического инструментария для наглядного представления связей [6]. Наиболее распространенным методом визуа-

лизации для ОСА является построение графа (сетевой карты), который показывает компоненты сети (сотрудников, департаменты, организации и т.д.) в виде узлов (вершин) и связующие их линии (связи или ребра).

Для визуализации результатов ОСА также могут применяться семантические карты, т.е. визуальную структуру связанных компонентов сети [18]. М. Хаспельмат (2003) описывает семантическую карту как геометрическое представление функций грамматической единицы (понятия), расположенных в «концептуальном/семантическом пространстве», соединенных линиями связей и организующих сеть [19]. Основная идея метода построения семантических карт и ключевое отличие от графов как визуализации сетей исследуемых понятий заключаются в том, что пространственное расположение различных понятий (например, организаций) отражает степень их сходства по близости расположения: чем ближе понятия расположены на карте, тем сильнее их связь [20]. При этом в зависимости от показателей, по которым рассчитывается такая связь, ее наличие и степень силы также могут быть отражены через ребра (соединительные линии между двумя отдельными понятиями). Таким образом, семантическое картирование позволяет выявлять, категориально структурировать, а также оценивать характер взаимосвязей между участниками сети.

Наиболее распространенный сценарий прикладного использования ОСА связан с исследованием внутренней среды организации или с анализом участников ближайшего окружения. В обоих случаях речь идет об анализе, как правило, ограниченного количества участников сети, о которых известно заранее, и его результаты часто связаны с мониторингом и диагностикой текущей ситуации, когда уже, вероятно, существуют некоторые предположения о наличии проблем или рисков. При этом зачастую не охватывается сегмент поисковых исследовательских задач, когда границы сетевого анализа задаются значительно

шире, а ожидаемыми результатами являются инсайты, имеющие прикладную ценность: выявление неизвестных ранее конкурентов, потенциальных партнеров и т.п.

Настоящее исследование ставит своей целью картирование потенциальных центров компетенций в области «Промышленные информационно-управляющие и телекоммуникационные системы» с отражением существующих связей между ними. В следующем разделе рассматриваются методологические возможности семантического анализа больших текстовых данных в системе iFORA для выявления организационных связей между центрами компетенций, а также описываются используемые данные. Затем приводятся результаты исследования, в рамках которого были выявлены связи участников сети с применением технологии анализа больших текстовых данных, после чего полученные результаты интерпретируются и обсуждаются.

1. Используемые данные и методологические возможности семантического анализа больших текстовых данных для выявления организационных связей центров компетенций

Современные программные решения позволяют сделать процесс ОСА более эффективным за счет возможностей анализа больших массивов данных и продвинутых способов визуализации. Подход к выявлению участников сети, реализованный в НИУ ВШЭ с использованием iFORA и рассматриваемый в настоящей работе, основан на принципах семантического поиска и машинного извлечения из релевантных документов «гипотетических» центров компетенций (центров компетенций, автоматически идентифицированных как релевантные исследуемому направлению, но не прошедших экспертную фильтрацию/валидацию). В настоящей статье под центрами компетенций понимаются крупные научные и исследовательские организации (включая вузы, технологические и промышленные компании), вносящие значимый вклад в развитие конкретного тематического направления.

Система iFORA¹ представляет собой инструмент интеллектуального анализа больших текстово-документных данных, основанный на углубленной обработке естественного языка с использованием искусственного интеллекта и машинного обучения. База данных iFORA насчитывает более 850 миллионов документов, включающих научные публикации, патенты, гранты и заявки на них, аналитические и стратегические документы, отраслевые профессиональные СМИ на английском, русском, китайском языках. Ключевое преимущество системы заключается в синтезе методов статистического, синтаксического и семантического анализа текстов для решения сложных аналитических задач. Система iFORA предоставляет модульную (пакетную) аналитику для различных исследовательских и бизнес-задач в сфере науки, технологий и инноваций. Анализ больших данных в системе iFORA в т. ч. предполагает извлечение из текстов терминов на основе расчета их текстовой статистики, а также последующее измерение семантической (векторной) близости этих терминов с использованием предобученных языковых моделей. Под термином понимается слово или словосочетание, синтаксически значимая конструкция, устойчиво употребляемая в большом массиве неструктурированных текстовых документов (а также в текстах научных публикаций, патентов, отраслевой аналитики и др.), которая является обозначением тематики — некоторого понятия (например, потенциального центра компетенций, как в границах настоящего исследования). Одной из функций системы является картирование центров и сетей компетенций, которое включает в себя выявление из больших текстовых данных перечней центров компетенций в т. ч. посредством анализа извлеченных метаданных документов, определение их специализации и др.

Также следует отметить, что возможности системы iFORA не заменяют, но дополняют и расширяют методологию ОА,

¹ URL: <https://issek.hse.ru/ifora/> (дата обращения: 03.06.2025).

которая тоже позволяет исследовать связи организаций и, безусловно, не может быть полностью воспроизведена с использованием анализа больших текстовых данных. Тем не менее преимущество использования системы iFORA заключается в том, что она позволяет формировать графы связей центров компетенций не просто исходя из совстречаемости их наименований в документах, но и отражая на графах кластеры центров компетенций в зависимости от «тематической близости» их деятельности (определяемой при помощи анализа закономерностей совместного употребления пар наименований центров компетенций в документах). Кроме того, задача извлечения наименований организаций из текстов для дальнейшего анализа их совместной упоминаемости сама по себе является нетривиальной и требует использования предобученных моделей машинного обучения в такой области обработки естественного языка, как распознавание именованных сущностей (Named Entity Recognition). Подобные модели включены в систему iFORA, которая также позволяет извлекать центры компетенций из метаданных загруженных в нее документов.

В системе iFORA используются три ключевых источника данных, соответствующие трем основным типам центров компетенций:

- научные публикации, соответствующие «научным» центрам компетенций (в большинстве случаев — университеты, научно-исследовательские институты, научные центры, лаборатории, кафедры и др., реже — промышленные компании, корпорации и др.);
- патенты, соответствующие «промышленным» центрам компетенций (R&D департаменты крупных организаций, осуществляющих промышленную разработку ИКТ и др.);
- отраслевые медиа-источники, соответствующие как университетам и научным организациям, так и «рыночным»

центрам компетенций (крупные организации — разработчики ИКТ-продуктов и решений для конечных пользователей; дополнительно сюда могут попадать организации-потребители, внедряющие ИКТ в собственную инфраструктуру).

В рамках проведенного исследования были задействованы три описанных типа источников информации, загруженных в iFORA, на английском языке за период 2011–2022 гг. в частности:

- база научных публикаций Microsoft Academic Graph (MAG);
- база международных патентов PATSTAT;
- агрегированная база зарубежных отраслевых профессиональных СМИ и рыночной аналитики.

В задачах анализа организационных связей используются, в частности, графы связей организаций, в т. ч. с тематической кластеризацией, получаемые на основе рассчитываемых количественных семантических метрик. Анализ текстовых данных и визуализация результатов такого анализа позволяют выявлять, в частности, ключевые центры компетенций, освещаемые в профессиональных источниках информации, их значимость в источниках и связи между отдельными центрами компетенций. Пример подобной визуализации представлен на рисунке 1, на котором изображен граф связей, где центры компетенций по направлению «Искусственный интеллект» отображаются в виде узлов разных цветов (в зависимости от принадлежности к кластерам по тематическим направлениям) с соответствующими подписями, отражающими наименования центров компетенций.

Размер узлов отражает значимость (интенсивность упоминания за рассматриваемый период) центра компетенций в источниках (анализируемой выборке документов), а связь между центрами компетенций, рассчитанная по совместной

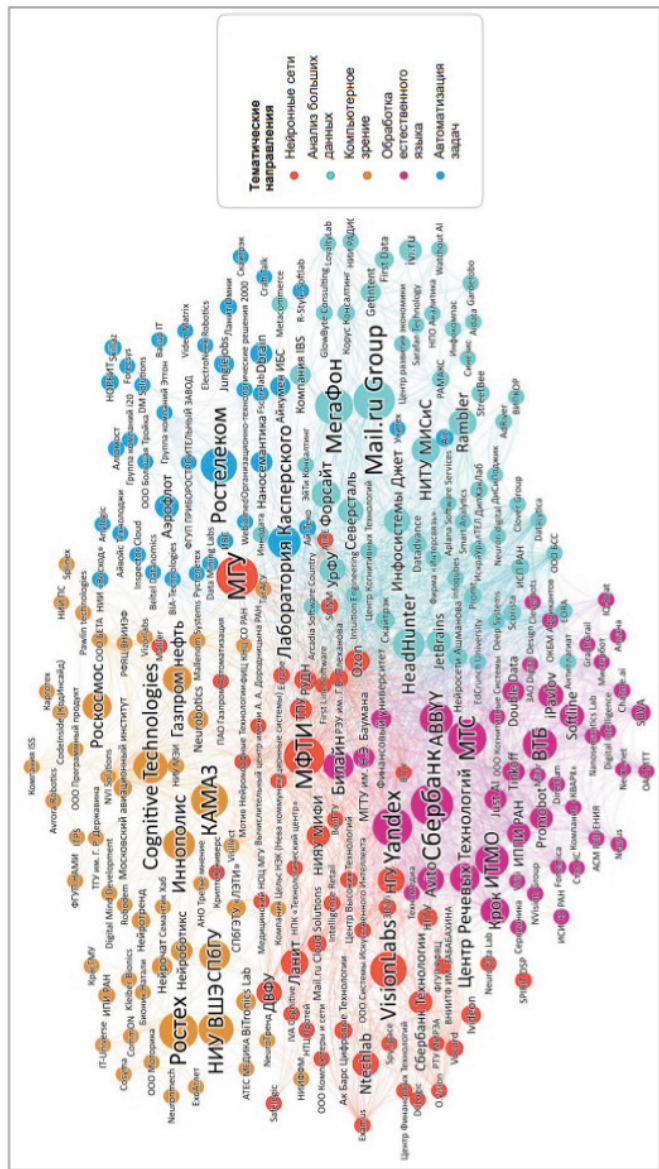


Рис. 1. Пример визуального представления графа связей центров компетенций по направлению «Искусственный интеллект»

упоминаемости в документах, показана с помощью линий (ребер) и степени удаленности соответствующих узлов друг от друга. Наиболее связанные центры компетенций входят в один тематический кластер по результатам анализа векторного представления данных центров компетенций и окрашиваются в один цвет. Такой анализ реализуется посредством специальной обученной дистрибутивной семантической модели, содержащей в себе научно-технические термины и показатели их смысловой близости друг с другом, которая оценивается не просто по совместной встречаемости терминов в одних и тех же документах, а с учетом сложных закономерностей их употребления в языке.

В качестве тематического направления, на примере которого была апробирована методология извлечения перечня ключевых центров компетенций с дальнейшим построением (визуализацией) организационных связей между ними, выбрано направление «Промышленные информационно-управляющие и телекоммуникационные системы». По поисковому условию, описывающему область «Промышленные информационно-управляющие и телекоммуникационные системы», машинным образом выявляются потенциальные («гипотетические») центры компетенций. «Гипотетические» центры компетенций в рамках исследуемого тематического направления могут извлекаться из текстов при помощи двух основных подходов: с использованием существующих метаданных документов — научных публикаций и патентов, содержащих, в частности, информацию о наименованиях аффилированных организаций, либо с применением распознавания именованных сущностей, выделенных системой iFORA автоматически при обработке заголовков и полных текстов документов — научных публикаций, патентов и отраслевой рыночной аналитики (таблица 1, рисунки 2–4).

Таблица 1

ДВА ПОДХОДА К ИЗВЛЕЧЕНИЮ ЦЕНТРОВ КОМПЕТЕНЦИЙ
ИЗ ДОКУМЕНТОВ: МЕТАДАННЫЕ И ИМЕНОВАННЫЕ
СУЩНОСТИ

	Метаданные документов	Выделенные именованные сущности (NER)
Научные публикации (MAG)	organisations	—
Патенты (PATSTAT)	assignee	—
Рыночная аналитика и профессиональные СМИ (собственная база документов системы iFORA)	—	phrases.en.all.ORG (на английском языке) и др.

«Гипотетическим» является тот центр компетенций, который был извлечен системой из больших текстовых данных, но не прошел экспертную фильтрацию/валидацию на предмет релевантности анализируемому направлению. Экспертная фильтрация/валидация автоматически извлеченных центров компетенций представляет собой экспертную работу в формате заочного опроса, а также в иных форматах (очных экспертных сессий) и включает в себя оценку актуальности центров компетенций для конкретного тематического направления, уточнение отдельных формулировок, перенос центров компетенций между сформированными кластерами и др. В рамках настоящего исследования фильтрация центров компетенций не производилась.

Результаты выявления «гипотетических» центров компетенций представляются в табличном виде, где каждый центр компетенций сопровождается набором текст-майнинговых метрик (текстовых статистик, таких как «встречаемость», «значимость», «динамичность», «ускорение», «осевой-год»; подробности раскрываются в следующем разделе), использование которых может обеспечить дополнительные преимущества аналитики, такие как

Рис. 2. Скриншот результатов обращения к базе документов iFORA и пример некоторых метаданных документов индекса «Научные публикации», в т.ч. содержащих сведения о центрах компетенций

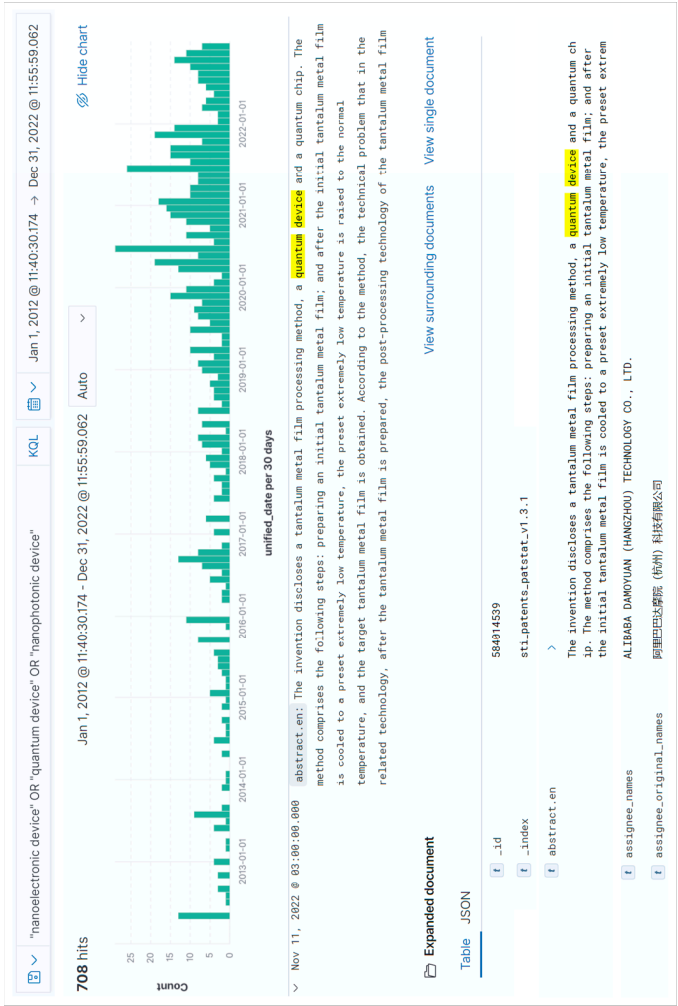


Рис. 3. Скриншот результатов обращения к базе документов iFORA и пример некоторых метаданных документов индекса «Патенты», в т.ч. содержащих сведения о центрах компетенций

Источник: система интеллектуального анализа больших данных iFORA.

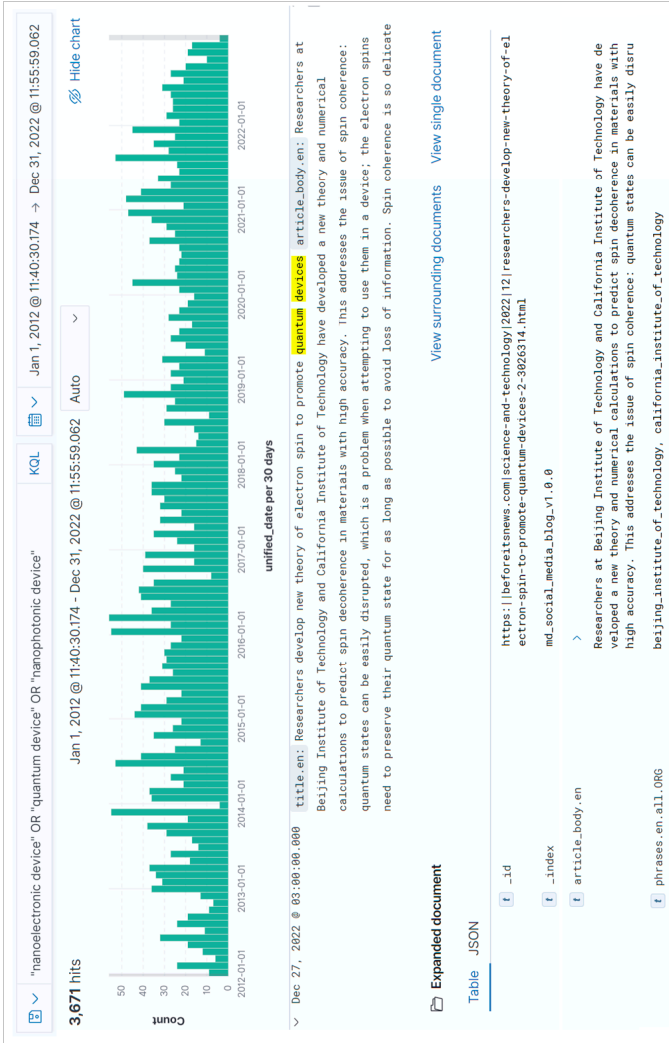


Рис. 4. Скриншот результатов обращения к базе документов iFORA и пример некоторых атрибутов документа «Рыночная аналитика и профессиональные СМИ», в т.ч. содержащих сведения о центрах компетенций»

сравнение центров компетенций друг с другом по объективным критериям, выявление не только крупных центров компетенций, но и наиболее важных из них, наиболее быстрорастущих и т.п. Такая фильтрация результатов может предшествовать процессу картирования организационных связей между выявленными центрами компетенций для того, чтобы полученные графы связей отражали только релевантные центры компетенций, исключая возможность попадания на визуализации «шумов» (ошибочно выделенных машинным образом центров компетенций) или просто невостребованных центров компетенций.

В таблице 2 приведена статистика извлеченных из документов центров компетенций. Всего по результатам обработки трех источников данных извлечено 2 268 наименований центров компетенций, из которых уникальными являются 1 645 (за исключением смысловых повторов). Отбор (экстракция) центров компетенций происходил с использованием экспертно-составленного поискового запроса (в части отбора релевантных документов) путем обращения к метаданным отобранных документов (в которых содержатся отдельные «поля» — наименования центров компетенций), а также при участии специализированных моделей машинного обучения, направленных на извлечение из текстов документов именованных сущностей типа «Организация» (таблица 1). Поисковый запрос по своей сути является аналогом семантического ядра исследуемой темы и предназначен для формирования границ предметной области. Он включает в себя термины, которые наиболее точно и полно описывают исследуемую тему (не являются слишком широкими или слишком узкими), соединенные друг с другом логическими операторами «AND» и «OR». Обращение к метаданным баз научных публикаций (MAG) и патентов (PATSTAT) позволило сформировать первые два списка «гипотетических» центров компетенций, а третий список был создан при помощи использования моделей машинного обучения, которые извлекают наименования центров компетенций из базы данных отраслевых СМИ.

Таблица 2
СТАТИСТИКА КОЛИЧЕСТВА ИЗВЛЕЧЕННЫХ ИЗ ДОКУМЕНТОВ РАЗЛИЧНЫХ ИСТОЧНИКОВ «ГИПОТЕТИЧЕСКИХ» ЦЕНТРОВ КОМПЕТЕНЦИЙ
 (по результатам использования специальной языковой модели системы iFORA, обученной на выявление именованных сущностей — наименований организаций, а также по результатам обработки метаданных проанализированных документов)

Направление	Поисковый запрос	Количество центров компетенций, в т. ч. в документах за 2012–2022 гг.		
		Научных публикаций (MAG)	Патентов (PATSTAT)	Профессиональных отраслевых СМИ
Промышленные информационно-управляющие и телекоммуникационные системы (в целом)	"information management system" OR "telecommunication system"	100	98	439
				Всего
				637

Продолжение табл. 2

Направление	Поисковый запрос	Количество центров компетенций, в т. ч. в документах за 2012–2022 гг.			
		Научных публикаций (MAG)	Патентов (PATSTAT)	Профес- сиональных отраслевых СМИ	Всего
функционально-сложные информационно- управляющие системы и процессы, методы и средства автоматизации	"functionally complex system" OR "information control system" OR "functionally complex process" OR "information control process"	100	100	74	274
интеллектуальные системы управления и принятия решений	"intelligent control system" OR "intelligent decision- making system"	100	96	93	289
интегрированные информационно- телекоммуникационные системы	"integrated information system" OR "integrated telecommunication system"	100	64	96	260

Окончание табл. 2

Направление	Поисковый запрос	Количество центров компетенций, в т. ч. в документах за 2012–2022 гг.			
		Научных публикаций (MAG)	Патентов (PATSTAT)	Профес- сиональных отраслевых СМИ	Всего
архитектура и программное обеспечение информационно- вычислительных комплексов	"information system architecture" OR "computing system architecture" OR "computing software"	100	93	92	285
информационная безопасность	"information security"	100	85	67	252
наноэлектронные, квантовые и нанофотонные	"nanoelectronic device" OR "quantum device" OR "nanophotonic device"	100	77	93	270

2. Результаты выявления организационных связей центров компетенций при помощи семантического анализа больших текстовых данных

Для наглядности при дальнейшем анализе был использован перечень центров компетенций, извлеченных для направления «Промышленные информационно-управляющие и телекоммуникационные системы». В целях фокусировки дальнейшего анализа на научных центрах компетенций в качестве источника для визуального представления — картирования графа связей в сети центров компетенций — выбрана база из 700 зарубежных научных публикаций MAG, откуда был извлечен топ-100 центров компетенций по семантическому показателю значимости (раскрывается подробнее ниже). При этом также доступны для построения графы связей и по другим типам источников (как отдельно, так и в совокупности), однако в рамках настоящей статьи они не рассматриваются. Для полноценного всестороннего исследования имеет смысл рассматривать набор таких графов связей, так как каждый из них может помочь сформировать гипотезы по различным типам и форматам взаимодействия между организациями.

На рисунке 5 представлен граф связей центров компетенций по направлению «Промышленные информационно-управляющие и телекоммуникационные системы» по данным зарубежных научных публикаций за 2012–2022 гг. Здесь визуализируется вся совокупность выявленных «гипотетических» центров компетенций по направлению «Промышленные информационно-управляющие и телекоммуникационные системы». Безусловно, приведенный граф связей может быть сложен для восприятия при первичном ознакомлении в силу своей плотности. Для повышения информативной ценности данной визуализации возможна ее декомпозиция на несколько графов, каждый из которых отражал бы центры компетенций, принадлежащие к одним и тем же кластерам. Однако подобное разделение исключает отображение связей между центрами

компетенций, принадлежащим к разным кластерам, что объясняет преимущество выбранного авторами способа визуального представления связей центров компетенций.

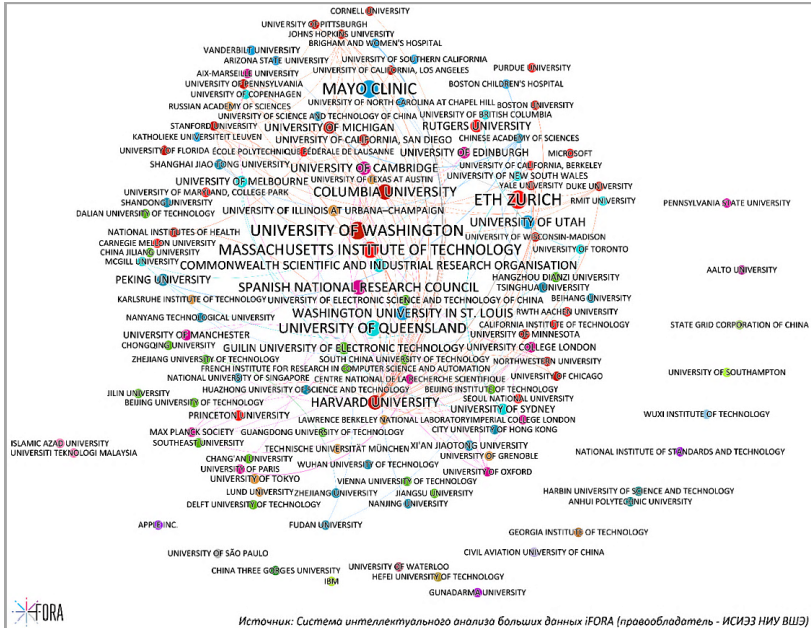


Рис. 5. Граф связей центров компетенций по направлению «Промышленные информационно-управляющие и телекоммуникационные системы» по данным зарубежных научных публикаций за 2012–2022 гг.

Уточним параметры полученного графа:

- **цвет узлов** отражает результаты математической кластеризации графа связей центров компетенций с использованием алгоритма Фрюхтермана-Рейнгольда, традиционно применяемого для визуального представления организационных

сетей; цвет указывает на близкие друг к другу центры компетенций, где показатель близости оценивается на основе паттернов совместного употребления названий центров компетенций в текстах;

- **расстояние между узлами** определяется аналогично с применением расчетов по алгоритму Фрюхтермана-Рейнгольда;
- **размер узла и его подписи** связан с показателем центральности (семантическим показателем, рассчитываемым в системе iFORA как квадратный корень от суммарной совстречаемости организации с другими организациями умноженной на квадратный корень от степени вершины графа; не путать с понятием «центральность» из теории графов и сетевого анализа), который, в свою очередь, рассматривается как эквивалент условной «важности» узла в графе;
- **цвет ребра** определяется цветами соединяемых узлов, т.е. может быть градиентным;
- **толщина ребра** пропорциональна показателю силы связи — совместной упоминаемости (совстречаемости) пары центров компетенций в текстах анализируемых документов; показатель нормируется и принимает значение до 1 (максимальная совстречаемость).

Граф отражает связи между топ-100 центрами компетенций. В приложении А (таблица 1) представлен полный перечень центров компетенций и их семантические метрики. Для удобства сопоставления визуальных и табличных сведений центры компетенций представлены в алфавитном порядке.

При семантическом анализе для центров компетенций рассчитывались следующие метрики:

- **встречаемость** — абсолютная частота упоминаемости центра компетенций в исследуемых документах за проанализированный период (абсолютная значимость); может быть только положительной;
- **значимость** — интенсивность упоминаемости центра ком-

петенций в документах за проанализированный период с учетом размера корпуса документов; чем выше значение, тем сильнее соответствующий центр компетенций представлен в исследуемом массиве документов; может быть только положительной;

- **динамичность** — среднегодовой темп прироста значимости центра компетенций за проанализированный период (может быть положительной и отрицательной);
- **ускорение** — темп роста динамичности центра компетенций за проанализированный период (отрицательные темпы роста или положительные); как правило, большинство центров компетенций в силу своей зрелости демонстрируют отрицательное ускорение, за исключением центров компетенций, обладающих рекордными темпами роста;
- **осевой год** — год, когда центр компетенций упоминался чаще чем за другие годы в границах проанализированного периода.

3. Интерпретация и обсуждение результатов

Можно выделить несколько сценариев интерпретации подобного графа связей центров компетенций:

Сценарий 1. Первичный анализ заключается в выделении отдельных организаций в научно-технологическом или рыночном ландшафте рассматриваемого направления. Такой анализ буквально отвечает на вопрос «Кто присутствует в исследуемом тематическом поле?», а семантические метрики могут внести вклад в оценку важности (масштабов присутствия) этих игроков.

Сценарий 2. Углубленный анализ подразумевает оценку наличия взаимосвязи между отдельными игроками и ее силы (интенсивности) с использованием рассчитанных количественных семантических метрик. В особенности такой анализ полезен, если организация знает своих конкурентов и, например, хочет понять их окружение.

Сценарий 3. Продвинутый анализ, включающий элементы экспертной интерпретации, связан с изучением качества этой выявленной взаимосвязи, по результатам которого можно получить информацию о том, какие типы взаимоотношений присутствуют и преобладают в исследуемой среде на основе оцененного текстового следа, его объема, тональности и других метрик. На этом этапе оцениваются конкретная роль и функционально-ролевые характеристики выявленных организаций. Данный сценарий одновременно обладает как наибольшей сложностью анализа, так и высокой ценностью получаемых результатов. Так, например, в зависимости от используемых для него источников данных можно выявить отдельных потенциальных соисполнителей по научным проектам (лаборатории с уникальным оборудованием и т.п.), поставщиков оборудования, субподрядчиков, клиентов, конкурентов и др. В широком понимании здесь можно говорить о перспективах реализации конкурентной разведки, поскольку рынок формируют его отдельные игроки, углубленный анализ которых (с точки зрения направлений деятельности, конкретных преимуществ, капитальных, кадровых и интеллектуальных ресурсов, компетенций, отношений с потребителями и т.п.) позволяет получить первичное представление о рынке и оценить собственное место на нем.

В рамках настоящей работы авторы постараются проиллюстрировать все три предложенных сценария работы с графом связей центров компетенций, которые в целом могут рассматриваться как этапы одного полноценного сценария. В рамках сценария 1 достаточно базового визуального анализа полученного графа связей (рисунок 2), можно воспользоваться таблицей семантических метрик (Приложение А, таблица 1).

В рамках сценария 2 для оценки наличия взаимосвязи между отдельными игроками и ее силы (интенсивности) используется семантический показатель встречаемости наименований центров компетенций. В рамках настоящего иссле-

дования в границах 100 выявленных центров компетенций зафиксировано 7 269 пар центров компетенций, между которыми существуют связи, граф (рисунок 2) визуализирует 100 наиболее сильных связей (т.е. изображено 100 ребер между 100 парами центров компетенций с наибольшим показателем совстречаемости; отбор топ-100 центров компетенций из MAG проводился по метрике значимости), а таблица 2 (Приложение А) демонстрирует топ-25 пар центров компетенций по направлению «Промышленные информационно-управляющие и телекоммуникационные системы», между которыми существует наиболее сильная связь.

Можно выделить три подхода к отбору пар центров компетенций для дальнейшего исследования:

- 1) на основе показателя силы связи между ними;
- 2) на основе семантических метрик отдельных организаций, например, значимости, динамичности, ускорения и др. (в таком случае исследователь сначала отбирает наиболее интересные для анализа центры компетенций, организации-лидеры, а далее оценивает связи выбранных центров компетенций с другими);
- 3) экспертным или поисковым методом, в первую очередь, для анализа собственного окружения или, например, окружения своих конкурентов, перечень которых известен заранее (прицельный сфокусированный анализ конкретных организаций и их связей с другими участниками рынка).

Рассмотрим несколько кейсов (топ-5 пар центров компетенций с наиболее сильной связью) и интерпретируем их в рамках иллюстрации сценария 3. Здесь следует отметить, что использование системы iFORA позволяет проводить анализ не только на уровне пар связей центров компетенций (от «сильных» к «слабым»), но и на уровне кластеров и при ранжировании пар связей организаций от «слабых» к «сильным». Выбранный в рамках настоящей статьи способ интерпретации результатов направлен на то, чтобы на конкретных и понятных примерах проиллюстри-

ровать валидность предлагаемой методологии, основанной на анализе больших текстовых данных.

1) Пара центров компетенций с наибольшей совстречаемостью: Harvard University и Brigham and Women's Hospital.

Поисковое исследование показывает, что Brigham and Women's Hospital — это учебная больница Гарвардской медицинской школы, поэтому сильная семантическая связь этих двух центров компетенций объяснима. При этом важно упомянуть, что с точки зрения метаданных, хранящихся в системе iFORA, данные центры компетенций (как и все организации, о которых далее идет речь) никак не связаны и являются по своей сути двумя отдельными сущностями. Медицинская тематика в области «Промышленные информационно-управляющие и телекоммуникационные системы» объясняется участием в специфических проектах на стыке медицины и IT [34]. Данный кейс иллюстрирует прямую дружественную аффилиацию. Роль центра компетенций в проекте — консультационная: BWH не занимается самостоятельной разработкой технологий, но руководит процессом и осуществляет контроль результатов с точки зрения т.н. образовательных эффектов. Безусловно, возможно выделить различные примеры сотрудничества двух указанных центров компетенций. Мы же лишь приводим отдельные кейсы, иллюстрирующие тот факт, что данные организации действительно связаны друг с другом, так как полноценная аналитическая работа по исследованию и подробному описанию различных аспектов аффилированности данных организаций (и центров компетенций из других кейсов) требует колоссальных ресурсов и выходит за рамки настоящей статьи.

2) Пара центров компетенций University of Science and Technology of China и Chinese Academy of Sciences.

Здесь наблюдается аналогичный тип связи, поскольку University of Science and Technology of China входит в структуру Chinese Academy of Sciences. Говоря о проектах, можно отметить

исследование по двухточечному распределению квантовых ключей на больших расстояниях, реализованное совместно со специалистами из Университета Цинхуа, Цзинаньского института квантовых технологий и Шанхайского института микросистем и информационных технологий (SIMIT), CAS [35]. Здесь говорится о дополнительных участниках проекта: исследовательская группа сотрудничала с Yangtze Optical Fiber и Cable Joint Stock Limited Company, которые, вероятно, выступили, производителями (разработчиками) «волокон со сверхмалыми потерями на основе технологии сердцевины из чистого кварца», благодаря чему стало возможно реализовать описанный эксперимент. Кейс иллюстрирует прямую дружественную аффилиацию University of Science and Technology of China и Chinese Academy of Sciences.

3) Пара центров компетенций University of Paris и Centre National de la Recherche Scientifique.

Два этих центра компетенций стали участниками проекта SNAP (Supernova Acceleration Probe) в интересах NASA, осуществляя совместную работу над специальной миссией по темной энергии [36]. В данном масштабном проекте принимали участие как промышленные партнеры (Ball Aerospace & Technologies, The Lockheed Martin Space Systems Company), так и академические (более 20 научных организаций и их отдельных подразделений), однако относительная семантическая связь указанных центров компетенций является менее выраженной, в связи с чем не отображена на сформированном графе. Кейс интересен тем, что позволяет не только выделить целую крупную сеть центров компетенций со схожим тематическим профилем и специализацией, но и определить таких участников в привязке к различным этапам/стадиям работ от фундаментальных исследований до прикладной реализации. Кейс не только иллюстрирует пример дружественной кооперации, но и демонстрирует широкий ландшафт потенциальных конкурентов в сегменте авиакосмических исследований.

4) Пара центров компетенций Massachusetts Institute of Technology и Harvard University.

Поисковое исследование совместных проектов этих двух центров компетенций позволяет сделать вывод о том, что они, в первую очередь, являются друг для друга потенциальными партнерами скорее в научно-образовательной сфере, нежели в прикладных направлениях. Есть сведения о совместных разработках в области мониторинга объектов в режиме реального времени [37], где не упоминаются дополнительные соучастники. Если говорить об общих исследованиях, то они позволяют более подробно прорисовать сеть связей. Так, например, MIT анонсировал выход новой книги, посвященной динамическому картированию городов [38]. Упоминается конкретная лаборатория MIT, которая занимала одну из центральных ролей в проекте — Senseable City Lab. Соавтором книги выступил сотрудник Harvard University's Graduate School of Design, приводится комментарий сотрудника University College London. Кейс демонстрирует небольшой тематический научный кластер центров компетенций в области урбанистики.

5) Пара центров компетенций Peking University и Chinese Academy of Sciences.

В качестве примера можно привести кейс совместной разработки устройства визуализации (сканера) магнитно-резонансной томографии исследователями из Шэньчжэньского института передовых технологий (SIAT) Китайской академии наук, которое впоследствии использовалось в Шэньчжэньской больнице Пекинского университета [39]. Исходя из открытых данных в кейсе, можно сделать вывод о том, что Пекинский университет выступил в качестве площадки для апробации и внедрения результатов исследований, в определенном смысле стал «потребителем» в общей цепочке трансфера разрабатываемой технологии (на это указывает и комментарий, содержащий предварительные оценки возможных экономических эффектов от внедрения

технологии). Кейс иллюстрирует пример производственной кооперации и промышленной интеграции, где лидерским ядром и носителем технологических знаний выступил SIAT. Китайская академия наук выступила одним из многочисленных партнеров, что указывает на потребность SIAT в сотрудничестве с государственными и академическими структурами для форсирования технологических разработок.

Представленные кейсы являются лишь некоторыми частными примерами того, как семантический анализ больших текстовых данных способен создавать пул гипотез, исследование и интерпретация которых может сформировать обширную базу знаний по игрокам рассматриваемого рыночного сегмента, их функционально-ролевым моделям и взаимосвязям. Такая информация может иметь конкретную практическую ценность при проведении исследований рынка, начиная от понимания тематической специализации организаций своей ближайшего окружения и заканчивая выстраиванием конкретных производственных цепочек продуктов и технологий.

Можно выделить ряд направлений развития описанного подхода для выявления центров компетенций из больших текстовых данных, его усиления и обогащения данными анализа текстовых документов. В дальнейшем целесообразно включить в анализ более широкий набор источников данных (объединив разные типы источников, проведя картирование всех 1 645 уникальных центров компетенций) с целью выявления различных типов организаций, связанных напрямую или косвенно с областью исследования, т.е. присутствующих на рынке: научных институтов, высших учебных заведений, конструкторских бюро, заводов, технологических и промышленных коммерческих организаций, государственных организаций и др.

Кроме того, перспективным видится выявление организационных связей в большем масштабе как среди мировых, так и среди российских центров компетенций. Система iFORA явля-

ется перспективным инструментом для анализа отечественных организаций по различным научно-техническим областям с использованием актуальных русскоязычных источников данных, включая научные публикации, патенты, документы рыночной аналитики и профессиональных СМИ.

Важно и развитие методологических подходов к картированию. Например, анализ динамики изменения количества центров компетенций (в рамках любой исследуемой области) возможно выполнять на основе различных периодов наблюдения: последние 10 лет, 3 года и отдельно за текущий год. Это позволит оценить робастность выявленных центров компетенций, определить устойчивые научно-технологические направления, идентифицировать новые быстрорастущие организации, а также отследить стагнирующие области исследований.

Заключение

В рамках настоящей статьи рассматривается технология анализа больших данных для выявления и анализа связей между отдельными центрами компетенций в области «Промышленные информационно-управляющие и телекоммуникационные системы». Возможности системы интеллектуального анализа больших данных iFORA, описанные в настоящей статье, позволяют эффективно решать задачи, которые традиционно требуют использования метода сетевого анализа, включая выявление отдельных элементов (участников) сети, определение связи между ними, а кроме этого — формирование гипотез по природе и характеру этих связей. Однако перспективным направлением исследований является и использование системы iFORA в сочетании с традиционными методами ОА, также предоставляющими методологический инструментарий для анализа и визуализации организационных связей, который возможно расширить за

счет обработки больших массивов текстовых данных в системе iFORA.

Система iFORA в качестве источников данных для анализа организаций использует научные публикации, патентные заявки, аналитические, новостные и иные открытые информационные ресурсы, что способствует улучшению качества работы методов и моделей автоматической обработки текстов в ходе картирования центров компетенций для выявления широкого перечня сведений, имеющих ценность для анализа и картирования организаций. В данной статье рассмотрен опыт формирования графа связей потенциальных центров компетенций области «Промышленные информационно-управляющие и телекоммуникационные системы», выявленных из англоязычных научных публикаций с использованием системы iFORA. Были предложены направления использования и развития описанной методологии, а также ее ключевое ограничение — необходимость экспертной фильтрации/валидации списков центров компетенций, автоматизированным образом извлеченных из документов, перед их непосредственным визуальным представлением на графе связей и его последующей интерпретацией. Кроме того, были сформированы и подтверждены перечни гипотез по существующим кооперациям, зафиксированные на основе семантически выявленных связей центров компетенций по исследуемому тематическому направлению. При этом подробные гипотезы в дальнейшем должны быть также проверены экспертным образом, что является следующим шагом типового аналитического сценария картирования центров компетенций.

Стоит также отметить важность постоянного расширения числа источников и охвата обрабатываемых данных. Доступ к полным текстам научных статей, патентов, грантов, отчетам о НИР и другим материалам позволит повысить качество обучаемых машинных моделей для выявления центров компетенций и надежного картирования связей между ними. Интерес пред-

ставляют не только российские источники информации, но и, в первую очередь, китайские, а также европейские, американские, включая международные патенты и гранты, материалы различных международных конференций, полные тексты научных публикаций в журналах высокого квартиля, рыночные аналитические отчеты. В результате становятся возможными выявление более разнообразных типов различных организаций и формирование первичных гипотез по видам взаимоотношений между ними. Например, при подключении баз стартапов станет доступен анализ их метаданных как отдельного типа организаций. Другим интересным источником для задач анализа организаций можно назвать площадки размещения данных о государственных закупках, в т. ч. данные электронных торгов, площадки размещения конкурсных процедур и т.п. с целью выявления взаимосвязей заказчиков и исполнителей (поставщиков). Обрабатывать такие разнообразные коллекции документов следует комплексно, например, с использованием онтологических переходников между языками, что даст возможность проводить исследования международного масштаба, к примеру, выявлять комплексные тематические кластеры центров компетенций.

ЛИТЕРАТУРА

1. *Hanneman R., Riddle M.* Introduction to Social Network Methods // Network. University of California. 2005. Vol. 46, № 7. 322 p.
2. *Морено Я. Л.* Кто останется в живых? Основы социометрии, групповой психотерапии и социодрамы / Пер. с англ. О. Шиловой. СПб.: Питер, 2024. 448 с. ISBN 978-5-4461-3936-1.
3. *Полякова А. Г., Колмаков В. В., Мирзабекова М. Ю.* Сетевой анализ организации социально-экономических процессов // Научно-технические ведомости СПбГПУ. Экономические науки. 2019, т. 12, № 3. С. 60–73. DOI: 10.18721/JE.12305. EDN: ZNXJVS.
4. *Wasserman S., Faust K.* Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press, 1994. 857 p. ISBN 9781139788618.
5. *Zhang M.* Social Network Analysis: History, Concepts, and Research //

Handbook of Social Network Technologies and Applications. Springer: Boston, MA, 2010. P. 3–21. ISBN 978-1-4419-7141-8. DOI: 10.1007/978-1-4419-7142-5_1.

6. *Freeman L.* The Development of Social Network Analysis // A Study in the Sociology of Science. 2004, № 1(687). P. 159–167. ISBN 9781594577147.

7. *Methot J. R., Zaman N., Shim H.* Social Network Analysis in Organizations // Oxford Research Encyclopedia of Business and Management, 2022. DOI: 10.1093/acrefore/9780190224851.013.228.

8. *Ramos V., Franco-Crespo A., González-Pérez L., Guerra Y., Ramos-Galarza C., Pazmiño P., Tejera E.* Analysis of Organizational Power Networks Through a Holistic Approach Using Consensus Strategies // Heliyon. 2019. № 5(2). P. 1–23. DOI: 10.1016/j.heliyon.2019.e01172.

9. *Allee V.* Value Network Analysis and Value Conversion of Tangible and Intangible Assets // Journal of Intellectual Capital. 2008. Vol. 9, № 1. P. 5–24. DOI: 10.1108/14691930810845777.

10. *Vogelsang J., Townsend M., Minahan M., Jamieson D., Vogel J., Viets A., Royal C., Valek L.* Handbook for Strategic HR: Best Practices in Organization Development from the OD Network. Amacom, 2012. 672 p. ISBN 9780814432495.

11. *Gursakal N., Oguzlar A., Berna Aydın Z., Tuzunturk S.* Measuring Trust in an Intra-Organisational Context Using Social Network Analysis // International Journal of Management and Enterprise Development, 2009. № 6(4). P. 494–512. DOI: 10.1504/IJMED.2009.024238.

12. *Tichy N., Fombrun Ch.* Network Analysis in Organizational Settings // Humanrelations. 1979. Vol. 32, № 11. P. 923–965. DOI: 10.1177/001872677903201103. EDN JMJHJ.

13. *Lock Lee L., Sica R.* Providing New Maps for Governance: Social and Organisational Network Analysis in Practice // Social Business Manifesto: [сайт]. 20.09.2012. URL: <https://sociallearning.it/category/organisational-network-analysis/> (дата обращения: 03.06.2025).

14. *Киселева Е. В., Крутцова М. Н., Приятелева Л. Г., Рудко А. М., Скворцова Л. И., Старцева С. Г.* Методы организационной диагностики в управлении персоналом: Учебно-методическое пособие. Вологда: Вологодский филиал РАНХиГС, 2016. 422 с. ISBN 978-5-906850-20-1. EDN: XITDEN.

15. *McDowell T., Horn H., Witkowski D.* Organizational Network Analysis Gain Insight, Drive Smart // Deloitte Touche Tohmatsu Limited. 2016. P. 2–4.

16. *Cross R.* Introduction to Organizational Network Analysis. 2004. URL: <https://gates.comm.virginia.edu/rcl3w/sna.htm> (дата обращения: 03.06.2025).

17. *Cross R.* Interpreting a Network Diagram. 2004. URL: <https://gates.comm.virginia.edu/rcl3w/sna03.htm> (дата обращения: 03.06.2025).

18. *Johnson D. D., Pittelman S. D., Heimlich J. E.* Semantic Mapping // The Reading Teacher. 1986. Vol. 39, № 8. P. 778–783.

19. *Haspelmath M.* The Geometry of Grammatical Meaning: Semantic Maps and Crosslinguistic Comparison // *The New Psychology of Language*. 2003, vol. 2. P. 211–242.
20. *Georgakopoulos T.* Semantic Maps // *Linguistics*. 2019. P. 1–23. DOI: 10.1093/OBO/9780199772810-0229.
21. *Mrvar A., Batagelj V.* Analysis and Visualization of Large Networks With Program Package Pajek // *Complex Adaptive Systems Modeling*. 2016, vol. 4, № 1. P. 1–8. DOI: 10.1186/s40294-016-0017-8. EDN: NFUCXD.
22. Сайт разработчика Pajek A. Mrvar [сайт]. URL: <http://mrvar.fdv.uni-lj.si/pajek/Symbols/symbols-examples.htm> (дата обращения: 03.06.2025).
23. Gephi [сайт]. URL: <https://gephi.org/> (дата обращения: 03.06.2025).
24. *Bastian M., Heymann S., Jacomy M.* Gephi: an Open Source Software for Exploring and Manipulating Networks // *Proceedings of the International AAAI Conference on Web and Social Media*. 2009. № 3(1). P. 361–362. DOI:10.1609/icwsm.v3i1.13937.
25. *Combe D., Largeron C., Egyed-Zsigmond E., G ry M.* A Comparative Study of Social Network Analysis Tools // *Web Intelligence & Virtual Enterprises*. 2010, vol. 2
26. *Carley K. M.* ORA: A Toolkit for Dynamic Network Analysis and Visualization // *Encyclopedia of Social Network Analysis and Mining*. Springer: New York, 2018. P. 1693–1702. DOI: 10.1007/978-1-4939-7131-2_309.
27. ORA-Lite [сайт]. URL: <http://www.casos.cs.cmu.edu/projects/ora/software.php> (дата обращения: 03.06.2025).
28. *Alhajj R., Rokne J.* Encyclopedia of Social Network Analysis and Mining. Springer Publishing Company, Incorporated, 2014. 2437 p. ISBN 978-1-4614-6169-2. DOI: 10.1007/978-1-4614-6170-8.
29. NetMiner [сайт]. URL: <http://www.netminer.com/product/features.do> (дата обращения: 03.06.2025).
30. *Smith M.A., Shneiderman B., Ilic-Frayling N., Rodrigues E.M., Barash V., Dunne C., Capone T., Perer A., Gleave E.* Analyzing (Social Media) Networks with NodeX // *Proceedings of the Fourth International Conference on Communities and Technologies*. 2009. P. 255–264. DOI:10.1145/1556460.1556497.
31. Social Media Research Foundation [сайт]. URL: <https://www.smrfoundation.org/> (дата обращения: 03.06.2025).
32. NodeXL [сайт]. URL: <https://nodexl.com/> (дата обращения: 03.06.2025).
33. The Social Media Research Foundation [сайт]. URL: <https://www.smrfoundation.org/nodexl/faq/> (дата обращения: 03.06.2025).
34. Jolly Good and Harvard University Hospital Collaborate on Medical VR - Full-scale Entry Into the U.S. Market // *PR Newswire* [сайт]. 22.03.2023. URL: <https://www.prnewswire.com/news-releases/jolly-good-and-harvard-university->

hospital-collaborate-on-medical-vr---full-scale-entry-into-the-us-market-301774703.html (дата обращения: 03.06.2025).

35. USTC Achieves Thousand-Kilometer Quantum Key Distribution // EurekAlert [сайт]. 13.01.2023. URL: <https://www.eurekalert.org/news-releases/992170> (дата обращения: 03.06.2025).

36. SNAP Wins NASA Support for Joint Dark Energy Mission // SPACENEWS [сайт]. 09.08.2006. URL: <https://spaceref.com/press-release/snap-wins-nasa-support-for-joint-dark-energy-mission/> (дата обращения: 03.06.2025).

37. MIT & Harvard's FAn System Reveal a Revolutionizing Real-Time Object Tracking // Cryptopolitan [сайт]. URL: <https://www.cryptopolitan.com/mit-reveal-real-time-object-tracking/> (дата обращения: 03.06.2025).

38. Mapping Cities in Motion // Massachusetts Institute of Technology [сайт]. 13.06.2023. URL: <https://news.mit.edu/2023/mapping-cities-motion-book-0613> (дата обращения: 03.06.2025).

39. China's Self-Developed MRI Machine Enters Mass Production // CGTN [сайт]. 24.07.2023. URL: <https://news.cgtn.com/news/2023-07-24/China-s-self-developed-MRI-machine-enters-mass-production-1IH76Ep3zck/index.html> (дата обращения: 03.06.2025).

Сведения об авторах

Логинова Ирина Владимировна

Заведующая отделом исследований больших данных Центра стратегической аналитики и больших данных Института статистических исследований и экономики знаний, Национального исследовательского университета «Высшая школа экономики».

Тел.: 8 (920) 036-70-96

SPIN РИНЦ: 2221-7707

ResearcherID: J-6034-2015

Пиекалнитс Анна Сергеевна

Ведущий эксперт отдела исследований больших данных Центра стратегической аналитики и больших данных Института статистических исследований и экономики знаний, Национального исследовательского университета «Высшая школа экономики».

Тел.: 8 (952) 065-11-71

Грозовский Фёдор Михайлович

Стажер-исследователь отдела исследований больших данных Центра стратегической аналитики и больших данных Института статистических исследований и экономики знаний, Национального исследовательского университета «Высшая школа экономики».

Тел.: 8 (925) 488-47-89

Приложение А

Таблица 1
 ПЕРЕЧЕНЬ ЦЕНТРОВ КОМПЕТЕНЦИЙ ПО НАПРАВЛЕНИЮ «ПРОМЫШЛЕННЫЕ
 ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ И ТЕЛЕКОММУНИКАЦИОННЫЕ СИСТЕМЫ»
 ПО ДАННЫМ ЗАРУБЕЖНЫХ НАУЧНЫХ ПУБЛИКАЦИЙ ЗА 2012-2022 ГГ.

Центр компетенций	Тип организации	Встречаемость (в источнике)	Значимость	Динамичность	Ускорение	Осевой год
Aalto University	университет	95679	0,010501	-0,01843	-0,08876	2016
Aix-Marseille University	университет	186542	0,020636	0,001081	-0,12229	2016
Arizona State University	университет	193148	0,021471	-0,01107	-0,08039	2016
Beihang University	университет	182713	0,020397	-0,00086	-0,0809	2016
Boston Children's Hospital	медицинский центр	436049	0,05025	0,008318	-0,07737	2016
Boston University	университет	206834	0,02318	-0,00918	-0,0718	2016
Brigham and Women's Hospital	университет (подразделение)	230428	0,026323	0,004849	-0,08384	2016
California Institute of Technology	университет	170994	0,018663	-0,02129	-0,07204	2015
Carnegie Mellon University	университет	159837	0,01757	-0,01437	-0,07755	2016

Продолжение табл. 1

Центр компетенций	Тип организации	Встречаемость (в источнике)	Значимость	Динамичность	Ускорение	Осевой год
Centre national de la recherche scientifique	научная организация	606339	0,065261	-0,0333	-0,07265	2015
Chinese Academy of Sciences	научная организация	1382284	0,157362	0,002422	-0,07938	2016
City University of Hong Kong	университет	115299	0,013254	0,006629	-0,07338	2016
Columbia University	университет	370255	0,041378	-0,01028	-0,0735	2016
Commonwealth Scientific and Industrial Research Organisation	научная организация	115793	0,012762	-0,02096	-0,06886	2015
Cornell University	университет	349001	0,039232	-0,00585	-0,07597	2016
Delft University of Technology	университет	168604	0,018696	-0,01399	-0,07245	2016
Duke University	университет	348021	0,039009	-0,00857	-0,0736	2016
École Polytechnique Fédérale de Lausanne	университет	183621	0,020126	-0,01947	-0,08007	2016
ETH Zurich	университет	248057	0,027518	-0,01049	-0,08338	2016

Продолжение табл. 1

Центр компетенций	Тип организации	Встречаемость (в источнике)	Значимость	Динамичность	Ускорение	Осевой год
French Institute for Research in Computer Science and Automation	научная организация	77997	0,008356	-0,03873	-0,06081	2015
Fudan University	университет	279747	0,032279	0,011049	-0,06683	2016
Georgia Institute of Technology	университет	203161	0,022501	-0,01253	-0,07187	2016
Gunadarma University	университет	50380	0,005606	-0,08142	0,056558	2011
Harvard University	университет	875720	0,098176	-0,0077	-0,07723	2016
Huazhong University of Science and Technology	университет	279405	0,032462	0,018196	-0,06692	2016
IBM	технологическая компания	110474	0,011902	-0,03217	-0,05849	2015
Imperial College London	университет	352220	0,039542	-0,00621	-0,08075	2016
Islamic Azad University	университет	343486	0,03868	0,003649	-0,10238	2016

Продолжение табл. 1

Центр компетенций	Тип организации	Встречаемость (в источнике)	Значимость	Динамичность	Ускорение	Осевой год
Johns Hopkins University	университет	473725	0,052518	-0,00958	-0,08282	2016
Karlsruhe Institute of Technology	университет	158642	0,01743	-0,01737	-0,08448	2016
Katholieke Universiteit Leuven	университет	379189	0,041969	-0,01581	-0,07508	2016
Lawrence Berkeley National Laboratory	научная организация	112670	0,012418	-0,01635	-0,08127	2016
Lund University	университет	214436	0,023785	-0,01734	-0,07512	2016
Massachusetts Institute of Technology	университет	368792	0,040389	-0,0184	-0,0799	2016
Max Planck Society	научная организация	685090	0,074089	-0,02497	-0,07953	2015
Mayo Clinic	медицинский центр	330115	0,037543	0,003358	-0,08964	2016
McGill University	университет	272178	0,030345	-0,01107	-0,07637	2016
Nanjing University	университет	382464	0,041733	-0,01149	-0,07393	2016
Nanyang Technological University	университет	232092	0,025871	-0,00679	-0,08865	2016

Продолжение табл. 1

Центр компетенций	Тип организации	Встречаемость (в источнике)	Значимость	Динамичность	Ускорение	Осевой год
National Institute of Standards and Technology	государственная организация	79606	0,008699	-0,01891	-0,08393	2015
National Institutes of Health	научная организация	415913	0,046111	-0,0183	-0,0707	2015
National University of Singapore	университет	306737	0,034758	-0,0006	-0,07885	2016
Northwestern University	университет	318836	0,035616	-0,01	-0,07443	2016
Peking University	университет	407895	0,046375	0,002082	-0,07119	2016
Pennsylvania State University	университет	295240	0,032887	-0,01293	-0,07045	2016
Princeton University	университет	187879	0,020666	-0,01553	-0,07187	2016
Purdue University	университет	283073	0,031019	-0,01859	-0,07516	2016
MIT University	университет	195129	0,022559	0,015881	-0,10019	2016
Russian Academy of Sciences	научная организация	622816	0,06788	-0,00844	-0,09734	2016
Rutgers University	университет	247025	0,027647	-0,00898	-0,07588	2016
RWTH Aachen University	университет	196044	0,021647	-0,01208	-0,09167	2016

Продолжение табл. 1

Центр компетенций	Тип организации	Встречаемость (в источнике)	Значимость	Динамичность	Ускорение	Осевой год
Seoul National University	университет	339512	0,037037	-0,02034	-0,07949	2016
Shanghai Jiao Tong University	университет	450764	0,051862	0,009927	-0,06955	2016
Spanish National Research Council	научная организация	421943	0,046313	-0,02269	-0,07253	2015
Stanford University	университет	519562	0,058116	-0,00582	-0,08042	2016
Technische Universität München	университет	253687	0,028301	-0,00323	-0,09368	2016
Tsinghua University	университет	417612	0,047292	0,004494	-0,08261	2016
Universiti Teknologi Malaysia	университет	117130	0,0126	-0,02788	-0,1062	2015
University College London	университет	415949	0,046731	-0,00444	-0,08727	2016
University of British Columbia	университет	349744	0,039422	-0,00692	-0,07708	2016
University of California, Berkeley	университет	342275	0,037876	-0,01391	-0,07439	2016

Продолжение табл. 1

Центр компетенций	Тип организации	Встречаемость (в источнике)	Значимость	Динамичность	Ускорение	Осевой год
University of California, Los Angeles	университет	414237	0,046319	-0,01193	-0,07386	2016
University of California, San Diego	университет	349945	0,039221	-0,00921	-0,07215	2016
University of Cambridge	университет	442391	0,049598	-0,00792	-0,0786	2016
University of Chicago	университет	241085	0,026723	-0,01452	-0,07868	2016
University of Copenhagen	университет	308224	0,03436	-0,0084	-0,0933	2016
University of Edinburgh	университет	278742	0,030955	-0,01289	-0,07919	2016
University of Florida	университет	313709	0,035328	-0,0064	-0,07006	2016
University of Grenoble	университет	82020	0,009686	0,043883	-0,18713	2017
University of Illinois at Urbana-Champaign	университет	301764	0,033276	-0,01834	-0,07341	2016

Продолжение табл. 1

Центр компетенций	Тип организации	Встречаемость (в источнике)	Значимость	Динамичность	Ускорение	Осевой год
University of Manchester	университет	281045	0,031304	-0,01765	-0,07021	2016
University of Maryland, College Park	университет	232703	0,025614	-0,01615	-0,07933	2016
University of Melbourne	университет	338800	0,038344	-0,00067	-0,08106	2016
University of Michigan	университет	541287	0,06044	-0,00935	-0,07648	2016
University of Minnesota	университет	383731	0,04273	-0,01239	-0,0762	2016
University of New South Wales	университет	308783	0,034716	-0,00406	-0,08757	2016
University of North Carolina at Chapel Hill	университет	303548	0,034153	-0,00646	-0,0776	2016
University of Oxford	университет	483413	0,05424	-0,00513	-0,08365	2016
University of Paris	университет	434781	0,049206	-0,00324	-0,08293	2016
University of Pennsylvania	университет	432226	0,048477	-0,00814	-0,07354	2016

Продолжение табл. 1

Центр компетенций	Тип организации	Встречаемость (в источнике)	Значимость	Динамичность	Ускорение	Осевой год
University of Pittsburgh	университет	331273	0,03718	-0,01091	-0,06839	2016
University of Queensland	университет	327466	0,03672	-0,01177	-0,07339	2016
University of Science and Technology of China	университет	211653	0,024519	0,025771	-0,08774	2016
University of Southampton	университет	183335	0,0202	-0,02021	-0,08179	2015
University of Southern California	университет	268952	0,030051	-0,01028	-0,07709	2016
University of Sydney	университет	383941	0,043286	0,000237	-0,08406	2016
University of Texas at Austin	университет	330457	0,036639	-0,01375	-0,07518	2016
University of Tokyo	университет	459900	0,050735	-0,01758	-0,0722	2016
University of Toronto	университет	494116	0,055818	-0,00405	-0,07766	2016
University of Utah	университет	210812	0,023675	-0,00667	-0,08382	2016

Окончание табл. 1

Центр компетенций	Тип организации	Встречаемость (в источнике)	Значимость	Динамичность	Ускорение	Осевой год
University of Washington	университет	473018	0,05291	-0,00985	-0,07587	2016
University of Waterloo	университет	150848	0,016861	-0,00829	-0,07347	2016
University of Wisconsin-Madison	университет	322934	0,035903	-0,01565	-0,07403	2016
Vanderbilt University	университет	215319	0,023496	-0,01784	-0,08066	2016
Vienna University of Technology	университет	93694	0,010186	-0,02275	-0,08085	2015
Washington University in St. Louis	университет	251333	0,0283	-0,00953	-0,06875	2016
Xi'an Jiaotong University	университет	234731	0,027194	0,023373	-0,08742	2016
Yale University	университет	341613	0,038564	-0,00492	-0,07499	2016

Источник: система интеллектуального анализа больших данных iFORA.

Таблица 2

ТОП-25 ПАР ЦЕНТРОВ КОМПЕТЕНЦИЙ ПО НАПРАВЛЕНИЮ
«ПРОМЫШЛЕННЫЕ ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ
И ТЕЛЕКОММУНИКАЦИОННЫЕ СИСТЕМЫ», МЕЖДУ
КОТОРЫМИ СУЩЕСТВУЕТ НАИБОЛЕЕ СИЛЬНАЯ СВЯЗЬ

Центр компетенций № 1	Центр компетенций № 2	Совместимость	Сила связи
HARVARD UNIVERSITY	BRIGHAM AND WOMEN'S HOSPITAL	89502	0,43
UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA	CHINESE ACADEMY OF SCIENCES	51500	0,25
UNIVERSITY OF PARIS	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	47437	0,23
MASSACHUSETTS INSTITUTE OF TECHNOLOGY	HARVARD UNIVERSITY	43306	0,21
PEKING UNIVERSITY	CHINESE ACADEMY OF SCIENCES	42647	0,20
STANFORD UNIVERSITY	HARVARD UNIVERSITY	36856	0,18
TSINGHUA UNIVERSITY	CHINESE ACADEMY OF SCIENCES	36574	0,17
UNIVERSITY OF SYDNEY	UNIVERSITY OF NEW SOUTH WALES	36426	0,17
UNIVERSITY OF CALIFORNIA, BERKELEY	LAWRENCE BERKELEY NATIONAL LABORATORY	34678	0,17
JOHNS HOPKINS UNIVERSITY	HARVARD UNIVERSITY	33401	0,16
UNIVERSITY OF PENNSYLVANIA	HARVARD UNIVERSITY	32733	0,16

Окончание табл. 2

Центр компетенций № 1	Центр компетенций № 2	Совместимость	Сила связи
UNIVERSITY OF MICHIGAN	HARVARD UNIVERSITY	32483	0,16
HARVARD UNIVERSITY	BOSTON UNIVERSITY	32224	0,15
UNIVERSITY OF WASHINGTON	HARVARD UNIVERSITY	31851	0,15
YALE UNIVERSITY	HARVARD UNIVERSITY	29214	0,14
UNIVERSITY OF OXFORD	UNIVERSITY OF CAMBRIDGE	28624	0,14
NANJING UNIVERSITY	CHINESE ACADEMY OF SCIENCES	27926	0,13
HARVARD UNIVERSITY	COLUMBIA UNIVERSITY	27099	0,13
ZHEJIANG UNIVERSITY OF TECHNOLOGY	ZHEJIANG UNIVERSITY	26734	0,13
HARVARD UNIVERSITY	DUKE UNIVERSITY	26492	0,13
NATIONAL INSTITUTES OF HEALTH	HARVARD UNIVERSITY	26333	0,13
SHANGHAI JIAO TONG UNIVERSITY	CHINESE ACADEMY OF SCIENCES	26307	0,13
UNIVERSITY OF OXFORD	UNIVERSITY COLLEGE LONDON	25955	0,12
UNIVERSITY OF TORONTO	UNIVERSITY OF BRITISH COLUMBIA	25790	0,12
UNIVERSITY OF OXFORD	IMPERIAL COLLEGE LONDON	25396	0,12

Источник: система интеллектуального анализа больших данных iFORA.

DOI: 10.19181/4m.2025.34.1.2

**BIG DATA ANALYSIS TECHNOLOGY FOR
ORGANIZATIONAL RELATIONSHIPS DETECTION:
CAPABILITIES OF THE IFORA SYSTEM**

Loginova Irina Vladimirovna

HSE University, Moscow, Russia

illoginova@hse.ru

ORCID: 0000-0002-3376-2728

Piekalnits Anna Sergeevna

HSE University, Moscow, Russia

apiekalnits@hse.ru

ORCID: 0000-0003-0585-5350

Grozovskiy Fedor Mikhailovich

HSE University, Moscow, Russia

fgrozovskiy@hse.ru

ORCID: 0009-0000-4342-7942

For citation: Loginova I.V., Piekalnits A.S., Grozovskiy F.M. Big Data Analysis Technology for Organizational Relationships Detection: Capabilities of the iFORA System. *Sotsiologiya: 4M (Sociology: methodology, methods, mathematical modeling)*, 2025, no. 60, p. 36-88. DOI: 10.19181/4m.2025.34.1.2

Abstract. The purpose of this article is to demonstrate approaches to analyzing connections between centers of competence using the iFORA system of intellectual analysis of large text-document data, illustrated by mapping the field of «Industrial Information Management and Telecommunication Systems». The results of the study include testing the approaches and methods implemented in the iFORA system on tasks traditionally addressed within the framework of organizational network analysis. The main result of the analysis is a network graph of connections between potential centers of competence in the field of «Industrial Information Management and Telecommunication Systems», identified from scientific publications using semantic analysis tools, reflecting

the connections between these centers. The graph is further given an interpretation by generating hypotheses about existing collaborations based on these connections. The article also provides suggestions for directions to utilize the research of centers of competence connections based on semantic analysis of large textual data arrays.

Keywords: data analysis, big data, semantic analysis, data mining, text mining, organizational network analysis, cooperation detection

Acknowledgments: the article was prepared within the framework of the Fundamental Research Program of the National Research University Higher School of Economics.

References

1. Hanneman R., Riddle M. Introduction to Social Network Methods, *Network. University of California*, 2005, vol. 46, no. 7. 322 p.
2. Moreno, J.L. *Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama* (transl., in Russian). St. Petersburg: Piter, 2024. 448 p.
3. Polyakova A.G., Kolmakov V.V., Mirzabekova M.Y. Network Analysis of Organization of Socio-Economic Processes (in Russian), *St. Petersburg Polytechnic University Journal: Economic Sciences*, 2019, vol. 12, no. 3, p. 60–73. DOI: 10.18721/JE.12305.
4. Wasserman S., Faust K. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press, 1994. 857 p.
5. Zhang M. “Social Network Analysis: History, Concepts, and Research”, in: *Handbook of Social Network Technologies and Applications*. Springer: Boston, MA, 2010, p. 3–21. DOI: 10.1007/978-1-4419-7142-5_1.
6. Freeman L. The Development of Social Network Analysis, *A Study in the Sociology of Science*, 2004, no. 1(687), p. 159–167.
7. Methot J.R., Zaman N., Shim H. Social Network Analysis in Organizations, *Oxford Research Encyclopedia of Business and Management*, 2022. DOI: 10.1093/acrefore/9780190224851.013.228.
8. Ramos V., Franco-Crespo A., González-Pérez L., Guerra Y., Ramos-Galarza C., Pazmiño P., Tejera E. Analysis of Organizational Power Networks Through a Holistic Approach Using Consensus Strategies, *Heliyon*, 2019, no. 5(2), p. 1-23. DOI: 10.1016/j.heliyon.2019.e01172.

9. Allee V. Value Network Analysis and Value Conversion of Tangible and Intangible Assets, *Journal of Intellectual Capital*, 2008, vol. 9, no. 1, p. 5–24. DOI: 10.1108/14691930810845777.
10. Vogelsang J., Townsend M., Minahan M., Jamieson D., Vogel J., Viets A., Royal C., Valek L. *Handbook for Strategic HR: Best Practices in Organization Development from the OD Network*. Amacom, 2012. 672 p.
11. Gursakal N., Oguzlar A., Berna Aydın Z., Tuzunturk S. Measuring Trust in an Intra-Organisational Context Using Social Network Analysis, *International Journal of Management and Enterprise Development*, 2009, no. 6(4), p. 494–512. DOI: 10.1504/IJMED.2009.024238.
12. Tichy N., Fombrun Ch. Network Analysis in Organizational Settings, *Human relations*, 1979, vol. 32, no. 11, p. 923–965. DOI: 10.1177/001872677903201103.
13. Lock Lee L., Sica R. “Providing New Maps for Governance: Social and Organisational Network Analysis in Practice”, in: *Social Business Manifesto* [site]. 20.09.2012. URL: <https://sociallearning.it/category/organisational-network-analysis/> (date of access: 03.06.2025).
14. Kiseleva E.V., Krutsova M.N., Priyateleva L.G., Rudko A.M., Skvortsova L.I., Startseva S.G. *Metody organizatsionnoi diagnostiki v upravlenii personalom: Uchebno-metodicheskoe posobie* [Methods of organizational diagnostics in personnel management: A teaching aid] (in Russian). Vologda: Vologodskii filial RANKhiGS Publ., 2016. 422 p.
15. McDowell T., Horn H., Witkowski D. Organizational Network Analysis Gain Insight, Drive Smart, *Deloitte Touche Tohmatsu Limited*, 2016, p. 2–4.
16. Cross R. *Introduction to Organizational Network Analysis*. 2004. URL: <https://gates.comm.virginia.edu/rlc3w/sna.htm> (date of access: 03.06.2025).
17. Cross R. *Interpreting a Network Diagram*. 2004. URL: <https://gates.comm.virginia.edu/rlc3w/sna03.htm> (date of access: 03.06.2025).
18. Johnson D.D., Pittelman S.D., Heimlich J.E. Semantic Mapping, *The Reading Teacher*, 1986, vol. 39, no. 8, p. 778–783.
19. Haspelmath M. The Geometry of Grammatical Meaning: Semantic Maps

- and Crosslinguistic Comparison, *The New Psychology of Language*, 2003, vol. 2, p. 211–242.
20. Georgakopoulos T. Semantic Maps, *Linguistics*, 2019, p. 1–23. DOI: 10.1093/OBO/9780199772810-0229.
 21. Mrvar A., Batagelj V. Analysis and Visualization of Large Networks With Program Package Pajek, *Complex Adaptive Systems Modeling*, 2016, vol. 4, no. 1, p. 1–8. DOI: 10.1186/s40294-016-0017-8.
 22. *Developer's website Pajek A. Mrvar* [site]. URL: <http://mrvar.fdv.uni-lj.si/pajek/Symbols/symbols-examples.htm> (date of access: 03.06.2025).
 23. *Gephi* [site]. URL: <https://gephi.org/> (date of access: 03.06.2025).
 24. Bastian M., Heymann S., Jacomy M. Gephi: an Open Source Software for Exploring and Manipulating Networks, *Proceedings of the International AAAI Conference on Web and Social Media*, 2009, no. 3(1), p. 361–362. DOI:10.1609/icwsm.v3i1.13937.
 25. Combe D., Largeron C., Egyed-Zsigmond E., G  ry M. A Comparative Study of Social Network Analysis Tools, *Web Intelligence & Virtual Enterprises*, 2010, vol. 2.
 26. Carley K.M. “ORA: A Toolkit for Dynamic Network Analysis and Visualization”, in: *Encyclopedia of Social Network Analysis and Mining*. Springer: New York, 2018, p. 1693–1702. DOI: 10.1007/978-1-4939-7131-2_309.
 27. *ORA-Lite* [site]. URL: <http://www.casos.cs.cmu.edu/projects/ora/software.php> (date of access: 03.06.2025).
 28. Alhajj R., Rokne J. *Encyclopedia of Social Network Analysis and Mining*. Springer Publishing Company, Incorporated, 2014. 2437 p. DOI: 10.1007/978-1-4614-6170-8.
 29. *NetMiner* [site]. URL: <http://www.netminer.com/product/features.do> (date of access: 03.06.2025).
 30. Smith M.A., Shneiderman B., ilic-Frayling N., Rodrigues E.M., Barash V., Dunne C., Capone T., Perer A., Gleave E. Analyzing (Social Media) Networks with NodeX, *Proceedings of the Fourth International Conference on Communities and Technologies*, 2009, p. 255–264. DOI:10.1145/1556460.1556497.
 31. *Social Media Research Foundation* [site]. URL: <https://www.smrfoundation.org/> (date of access: 03.06.2025).
 32. *NodeXL* [site]. URL: <https://nodexl.com/> (date of access: 03.06.2025).

33. *The Social Media Research Foundation* [site]. URL: <https://www.smrfoundation.org/nodex1/faq/> (date of access: 03.06.2025).
34. Jolly Good and Harvard University Hospital Collaborate on Medical VR - Full-scale Entry Into the U.S. Market, in: *PR Newswire* [site]. 22.03.2023. URL: <https://www.prnewswire.com/news-releases/jolly-good-and-harvard-university-hospital-collaborate-on-medical-vr---full-scale-entry-into-the-us-market-301774703.html> (date of access: 03.06.2025).
35. USTC Achieves Thousand-Kilometer Quantum Key Distribution, in: *EurekAlert* [site]. 13.01.2023. URL: <https://www.eurekalert.org/news-releases/992170> (date of access: 03.06.2025).
36. SNAP Wins NASA Support for Joint Dark Energy Mission, in: *SPACENEWS* [site]. 09.08.2006. URL: <https://spaceref.com/press-release/snap-wins-nasa-support-for-joint-dark-energy-mission/> (date of access: 03.06.2025).
37. MIT & Harvard's FAN System Reveal a Revolutionizing Real-Time Object Tracking, in: *Cryptopolitan* [site]. URL: <https://www.cryptopolitan.com/mit-reveal-real-time-object-tracking/> (date of access: 03.06.2025).
38. Mapping Cities in Motion, in: *Massachusetts Institute of Technology* [site]. 13.06.2023. URL: <https://news.mit.edu/2023/mapping-cities-motion-book-0613> (date of access: 03.06.2025).
39. China's Self-Developed MRI Machine Enters Mass Production, in: *CGTN* [site]. 24.07.2023. URL: <https://news.cgtn.com/news/2023-07-24/China-s-self-developed-MRI-machine-enters-mass-production-1lH76Ep3zck/index.html> (date of access: 03.06.2025).

Information about the authors

Irina Vladimirovna Loginova

Head of Big Data Research Unit in the Center for Strategic Analytics
and Big Data, ISSEK, NRU HSE

SPIN-code: 2221-7707

ResearcherID: J-6034-2015

Anna Sergeevna Piekalnits

Leading expert of Big Data Research Unit in the Center for Strategic
Analytics and Big Data, ISSEK, NRU HSE

Fedor Mikhailovich Grozovskiy

Research Assistant of Big Data Research Unit in the Center for
Strategic Analytics and Big Data, ISSEK, NRU HSE