

Применение технологий извлечения, экспликации и анализа мультимодального контекстного знания при исследовании терминологической базы междисциплинарного направления «Информатизация научной деятельности»

О.В. Кононова, Д.Е. Прокудин

Аннотация— Цифровые трансформации современного информационного общества, хотя зачастую и опережают аналогичные процессы в науке, заставляют исследователей совершенствовать используемые технологии, заимствовать применяемые подходы из других сфер деятельности и смежных предметных областей, что в свою очередь ведет к развитию аналитического программного инструментария для научных целей. Особенно ярко эти явления и тенденции проявляются в развивающихся междисциплинарных научных направлениях, не представимых сегодня без ИКТ. Исследования по тематике «Информатизация научной деятельности» пока воспринимаются большинством ученых как недостаточно целостное, не полностью сформированное междисциплинарное направление. Предлагаемые результаты и выводы направлены на восполнение имеющихся лагун за счет выделения и изучения терминологического ядра направления. Разработанный авторами подход (синтетический метод) к исследованиям развития тематик и понятийно-терминологического аппарата междисциплинарных научных направлений позволяет выявить основные термин-концепты, участвующие в формировании тезауруса предметной области, основные группы текстовой и нетекстовой модальности данных, особенности мультимодальных научных документов и материалов. Сделан вывод интегрального характера контекстного знания, заключающемся во взаимосвязи информации текстовой и нетекстовой модальности, которую необходимо учитывать при проведении научных исследований контекстного знания.

Ключевые слова— контекстное знание, контекстный поиск, мультимодальность, научная деятельность, информатизация, междисциплинарные научные направления, терминологическая база, тезаурус, термин-концепт, Voyant-Tools, Sketch Engine.

Статья получена 30 ноября 2020.

Кононова Ольга Витальевна, Университет ИТМО, канд. экон. наук, доцент, (e-mail: koноnol@yandex.ru).

Прокудин Дмитрий Евгеньевич, Санкт-Петербургский государственный университет, докт. филос. наук, доцент; Университет ИТМО (e-mail: hogben.young@gmail.com).

Статья подготовлена по итогам выступления на Всероссийской конференции «Научный сервис в сети Интернет» (Абрау-2020) и дополнена новыми результатами.

I. ВВЕДЕНИЕ

Современное цивилизационное развитие характеризуется тотальными процессами информатизации всех областей человеческой деятельности, всех пространств его существования. Информатизация основывается на всестороннем применении информационно-коммуникационных технологий (ИКТ), которые являются продуктом научно-технического прогресса. Как порождение научной деятельности ИКТ находят своё применение и в самой науке, в научной деятельности, при проведении научных исследований. Поэтому информатизация характерна и для этой области человеческой деятельности. Однако, информатизация науки является внутренней потребностью научного сообщества и развивается в соответствии с логикой внутреннего развития научных исследований и потребностями учёных и исследователей. В связи с этим, например, на государственном уровне в России нет единой концепции информатизации научной деятельности (в отличие от системы образования, в которой на государственном уровне в разное время разрабатывались и реализовывались с разной степенью успешности концепции и программы информатизации образования). Об отдельных направлениях информатизации научных исследований упоминается только на уровне отдельных учреждений высшего образования [15, 24]. В подобного рода документах приводятся только общие слова о необходимости информатизации научных исследований без конкретизации и обоснования применения тех или иных методов и подходов организации и проведения научных исследований на основе применения ИКТ. При этом информатизацию научной деятельности необходимо разделять на два независимых направления:

- 1) информатизация научных исследований;
- 2) информатизация процессов управления организации научной деятельности в обществе.

Для решения задач проводимого исследования рассматривается первое направление как наиболее значимое. Для более эффективной организации научной

деятельности и. Цели и результаты исследования терминологической базы междисциплинарного направления «Информатизация научной деятельности» с выявлением особенностей и специфики информатизации как отдельных видов научной деятельности, так и ее отдельных тематических направлений соответствуют потребностям науки, научного сообщества и направлению «фундаментальные исследования, обусловленные внутренней логикой развития науки, обеспечивающие готовность страны к большим вызовам, еще не проявившимся и не получившим широкого общественного признания, возможность своевременной оценки рисков, обусловленных научно-технологическим развитием» из Стратегии научно-технологического развития Российской Федерации, что обуславливает их актуальность.

Развиваемая авторская методика поиска, извлечения, уточнения, экспликации, анализа и представления контекстного знания на основе комплексного использования ИКТ апробируется и используется при проведении исследований развития тематик и понятийно-терминологического аппарата междисциплинарных научных направлений. Этот подход к науковедческим исследованиям назван авторами «синтетическим методом». Он предполагает инвариантность и независимость методик от применяемого инструментария. Поэтому исследователи могут в своих целях использовать доступные им программные продукты и информационные системы. Этот подход показал свою эффективность при исследовании развития терминологической базы междисциплинарных направлений научных исследований «Цифровая экономика: электронное государственное управление и умные технологии» [12] и «Социально-этические аспекты цифровой экономики: геймификация в социальной и научно-образовательной сферах» [14].

В ходе исследования выявляется тематика основных направлений «Информатизации научной деятельности» и формируется его терминологическое ядро. Далее на основе экспликации и анализа контекстного знания уточняется терминологический ландшафт исследуемого направления, выявляются семантические группы и контексты использования терминов. Основными задачами исследования являются:

- выявление основных тематических направлений;
- определение основных источников исследования;
- формирование терминологического ядра исследуемого направления;
- уточнение терминологического ландшафта исследуемого направления;
- выявление контекстов использования терминов;
- выявления связи контекстов текстовой и нетекстовой модальности, используемых в научных исследованиях.

Исследование проводилось на массивах информации, полученных из различных цифровых источников, представляющих собой научные публикации за последние 10 лет. В качестве инструментов экспликации и анализа контекстного знания используются системы

Voyant-Tools и Sketch Engine.

II. МЕЖДИСЦИПЛИНАРНЫЙ ХАРАКТЕР ОБЛАСТИ НАУЧНЫХ ИССЛЕДОВАНИЙ «ИНФОРМАТИЗАЦИЯ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ»

Анализируя тенденции последнего десятилетия внедрения информационно-коммуникационных технологий с практику проведения научных исследований можно с уверенностью говорить о междисциплинарности исследуемого направления. При этом достаточно уверенно возможно выявить основные направления его развития, что позволит сформировать ядро терминологической базы, а также использовать элементы этого ядра для поиска и экспликации контекстного знания в целях дальнейшего исследования и формирования тезауруса. Междисциплинарность рассматриваемого направления определяется двумя основными факторами.

1. Методы исследований в различных научных областях предполагают использование ИКТ как инструмента проведения исследования. При этом в рамках некоторых направлений научных исследований такие методы сформировались в отдельные направления, которые зачастую также являются междисциплинарными, так как применяются представителями различных научных областей:

- вычислительная лингвистика и компьютерные онтологии (в прикладной лингвистике и языкознании);
- суперкомпьютерные технологии, параллельные вычисления (в прикладной математике, физике, астрономии, биоинформатике и др.) [7, 8, 10, 11, 23, 27];
- распределённые вычисления (применяется в различных научных областях) [1, 20, 28];
- цифровая гуманитаристика (гуманитарные науки) [4, 17, 25, 26];
- историческая информатика и цифровая история как самостоятельное ответвление цифровой гуманитаристики [2, 3, 5, 19];
- компьютерное моделирование (широко применяется в различных научных областях) [6, 16, 18, 29].

2. Помимо самих исследований научная деятельность представляет собой устоявшиеся виды и формы деятельности, которые обеспечивают организацию, проведение научных исследований, а также представление их результатов. К основным видам деятельности следует отнести:

- информационно-поисковый, в рамках которого происходит информационное обеспечение научных исследований;
- подготовка научных текстов, т.е. оформление результатов исследований;
- обнародование результатов научных исследований через публикации и выступления на публичных научных мероприятиях;
- научная коммуникация, которая обеспечивает: распространение научной информации в обществе, установление академических связей, проведение совместных научных исследований и многое другое.

Необходимо отметить, что использование информационно-коммуникационных технологий для поддержки организации и проведения научной деятельности инвариантно по отношению как к научной области, так и к конкретному исследованию. Эти технологии являются стандартными и в конкретном случае связаны только с выбором определённого аппаратного и программного обеспечения, которые являются доступными исследователям.

III. ОТБОР РЕСУРСОВ ДЛЯ ПРОВЕДЕНИЯ ИССЛЕДОВАНИЯ

Для выявления тематических направлений и ядра терминологической базы направления междисциплинарных исследований «Информатизация научной деятельности» был произведён отбор основных источников научной информации, в которых рассматриваются различные аспекты применения ИКТ в научной деятельности. На данном этапе исследования рассматриваются русскоязычные источники. Поэтому как основные источники использовались российские научные периодические и сериальные издания, а также материалы научных конференций. Этот массив наиболее полно представлен на цифровой платформе Научной электронной библиотеки (НЭБ, <http://elibrary.ru>). В поле нашего зрения также попали электронный репозиторий российских журналов открытого доступа Киберленинка (<http://cyberleninka.ru>) и специализированная информационно-поисковая система Академия Гугл (<https://scholar.google.ru>). Однако от их использования мы отказались, так как в этих электронных системах информация представлена в неструктурированном виде, отсутствуют удобные расширенные возможности поиска и аналитические инструменты.

В отличие от них в НЭБ реализованы:

- расширенный поиск с возможностью фильтрации;
- формирование тематических подборок публикаций;
- аналитические инструменты анализа тематических подборок, а также отдельных периодических и сериальных изданий.

Основываясь на анализе предметного поля исследования и собственном опыте использования информационно-коммуникационных технологий в различных междисциплинарных проектах для анализа были отобраны следующие издания:

– Научный сервис в сети Интернет (https://www.elibrary.ru/title_about_new.asp?id=68214) – сериальное издание, публикующее материалы одноимённой всероссийской научной конференции. В настоящее время в НЭБ представлены выпуски за 2018 и 2019 годы.

– Информационное общество: образование, наука, культура и технологии будущего (https://www.elibrary.ru/title_about.asp?id=66763) –

сериальное издание, публикующее материалы международной объединённой конференции «Интернет и современное общество». В настоящее время в НЭБ представлены выпуски за 2017, 2018 и 2019 годы.

– Компьютерная лингвистика и вычислительные онтологии (https://www.elibrary.ru/title_about.asp?id=66764) –

сериальное издание, публикующее материалы международной объединённой конференции «Интернет и современное общество». В настоящее время в НЭБ представлены выпуски за 2017, 2018 и 2019 годы.

– Государство и граждане в электронной среде (https://www.elibrary.ru/title_about.asp?id=66765) –

сериальное издание, публикующее материалы международной объединённой конференции «Интернет и современное общество». В настоящее время в НЭБ представлены выпуски за 2017, 2018 и 2019 годы.

– International Journal of Open Information Technologies, INJOIT (https://www.elibrary.ru/title_about.asp?id=41045) –

электронный научный журнал открытого доступа, публикующий статьи в области информационных технологий, многие из которых отражают результаты исследований из различных научных областей с использованием ИКТ. В НЭБ представлены номера начиная с 2013 года.

– Научная визуализация (https://www.elibrary.ru/title_about.asp?id=30470) –

электронный научный журнал, освещающий различные аспекты разработки и практического использования инструментальных средств и методов компьютерной визуализации научных данных. В НЭБ представлены номера начиная с 2009 года.

– Videонаука (https://www.elibrary.ru/title_about.asp?id=58109) –

электронный научный журнал, использующий компьютерное видео для распространения результатов научных исследований. В НЭБ представлены номера начиная с 2016 года.

– Электронные библиотеки (https://www.elibrary.ru/title_about_new.asp?id=1829) –

электронный научный журнал, который посвящён оперативному отражению новых результатов теории и практики использования в России и за рубежом электронных библиотек, которые, в том числе, выполняют важную роль в информационном обеспечении научных исследований. В НЭБ представлены номера начиная с 1998 года;

– Гуманитарная информатика (https://www.elibrary.ru/title_about.asp?id=9521) –

электронный научный журнал, посвящённый развитию междисциплинарных оснований гуманитарных исследований и обмену практическим опытом их проведения, в том числе на основе использования ИКТ. В НЭБ представлены номера начиная с 2004 года;

– Историческая информатика. Информационные технологии и математические методы в исторических исследованиях и образовании (https://www.elibrary.ru/title_about.asp?id=37732) –

электронный научный журнал, публикующий результаты научных исследований, статьи и обзоры, посвящённые информационным и математическим методам и технологиям исторического исследования. В НЭБ представлены номера начиная с 2012 года.

Эти массивы публикаций были дополнены следующими источниками:

– Сборники научных трудов научно-практического семинара «Информационное обеспечение науки: новые

технологии» (http://www.benran.ru/SEM/Sem3_83.html), который отражает развитие использования ИКТ в информационной поддержке научных исследований (с 1985 года).

– Сборники тезисов докладов (с 2011 года) и статей (с 2011 по 2016 годы) международной объединённой конференции «Интернет и современное общество» (<http://ojs.itmo.ru/index.php/IMS>).

– Сборники научных трудов Всероссийской научной конференции «Научный сервис в сети Интернет» за 2015, 2016 и 2017 годы (<http://agora.guru.ru/display.php?conf=abrau2020&page=ite m1>).

В НЭБ разработан API (Application Programming Interface), предназначенный для осуществления автоматических запросов и извлечения информации из базы данных РИНЦ. Однако, он является платным. Поэтому в рамках исследования использовался бесплатный функционал расширенного поиска, экспликации и анализа библиографической информации.

IV. ВЫЯВЛЕНИЕ ОСНОВНЫХ ТЕМАТИЧЕСКИХ НАПРАВЛЕНИЙ «ИНФОРМАТИЗАЦИИ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ»

Для выявления основных тематических направлений развития направления междисциплинарных исследований «Информатизация научной деятельности» сначала были получены данные по всем научным публикациям из НЭБ с использованием запроса ("информатизация науки" or "информатизация научной деятельности" and not образование). Было получено всего 33 публикации, что говорит о том, что понятия «информатизация науки» и «информатизация научной деятельности» крайне редко используются в научном дискурсе, а само научное направление «Информатизация научной деятельности» ещё далеко от целостного формирования. В связи с этим были проанализированы только подборки отобранных на начальном этапе периодических и сериальных изданий. Для этого в инструменте «Анализ публикационной деятельности журнала» по каждому изданию были получены статистические отчёты «Распределение публикаций по тематике». Из рассмотрения были удалены следующие тематики: «Информатика», «Автоматика. Вычислительная техника» и «Кибернетика». Это объясняется тем, что ИКТ в контексте исследуемого направления рассматриваются нами в контексте их использования в качестве инструментов исследования в различных научных направлениях и областях. Также не рассматривалось тематическое направление «Народное образование. Педагогика», так как в основном в рамках этого направления рассматриваются вопросы информатизации образования, не являющейся предметом нашего исследования. По всем изданиям были обобщены данные и выделены первые 10 по популярности тематических направлений, что вполне характеризует информатизацию научной деятельности в России (рис. 1).



Рис. 1. Тематическое распределение публикаций из отобранных периодических и сериальных изданий (топ-10)

Помимо отмеченных на рисунке, информационно-коммуникационные технологии применяются в исследованиях по следующим направлениям: «Государство и право. Юридические науки», «Биология», «Религия. Атеизм», «Сельское и лесное хозяйство», «Геофизика», «Машиностроение», «Науковедение», «Химия», «Геология» и некоторые другие. В целом, анализ полученных данных позволяет подтвердить междисциплинарный характер исследуемого направления научных исследований. При этом информатизация широко охватывает различные направления и области научных исследований.

V. ФОРМИРОВАНИЕ ТЕРМИНОЛОГИЧЕСКОГО ЯДРА «ИНФОРМАТИЗАЦИИ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ»

Следующим этапом исследования является выявление терминологического ядра рассматриваемого направления научных исследований. Для решения этой задачи при помощи аналитических инструментов НЭБ на отобранном массиве релевантных публикаций были проанализированы статистические отчёты «Распределение публикаций по ключевым словам». На основе анализа в терминологическое ядро были отобраны ключевые слова с числом упоминаний в публикациях не ниже 10 (при максимальном числе 88). В ядро не были включены ключевые слова, не релевантные исследуемому направлению научных исследований (например, «образование», «информационное общество»).

Таким образом, в терминологическое ядро направления междисциплинарных исследований «Информатизация научной деятельности» были включены следующие опорные термин-концепты: «визуализация» («научная визуализация»), «визуальная аналитика» («визуальный анализ»), «анализ данных», «базы данных», «онтологии», «блокчейн», «виртуальная реконструкция», «интернет», «информационные технологии» («ИКТ»), «информационные системы», «моделирование», «прикладная программа» («программа»), «визуализация данных», «географические информационные системы» («ГИС»).

Кроме этого, в соответствии с тематическими направлениями, выявленными на предыдущем этапе, кандидатами на включение в терминологическое ядро можно считать: «виртуальная реальность»,

«искусственный интеллект», «3d-модель» («3d-моделирование»), «графическая интерпретация», «машинное обучение», «многомерные данные», «цифровая гуманитаристика» («цифровые гуманитарные науки»), «контент-анализ», «орепл», «издательские системы», «коммуникация», «анализ программ», «анализ текстов», «параллельное программирование», «электронные библиотеки», «электронный научный журнал». Решение о включении их в ядро может быть принято по результатам контекстного анализа, который будет проведён на следующем этапе исследования. Но, тем не менее, эти термин-концепты будут включены в терминологическую базу исследуемого направления исследований.

Также были проанализированы подборки публикаций из НЭБ, сформированные по релевантным для исследуемого направления научных исследований запросам: «информатизация науки» and not образование (178 публикаций), «информатизация науки» or «информатизация научной деятельности» and not образование (20 публикаций), «цифровая гуманитаристика» or «цифровые гуманитарные науки» or «digital humanities» (476 публикаций), историческая информатика (41 публикация).

На основе анализа распределения ключевых слов в подборках были отобраны термины ядра («информатизация науки», «информатизация научной деятельности», «цифровизация», «информационное пространство научных исследований», «историческая информатика», «гуманитарная информатика», «цифровая история»), термин-концепты терминологической базы («аккумуляция и распространение результатов научных исследований», «большие данные», «междисциплинарность», «интернет-исследования», «информационная среда», «наукометрия»), а также термины, являющиеся синонимами уже отобранных («информационное пространство», «цифровые технологии», «цифровые ресурсы», «дигитализация», «цифровая среда»).

VI. ИССЛЕДОВАНИЕ ТЕРМИНОЛОГИЧЕСКОГО ЛАНДШАФТА «ИНФОРМАТИЗАЦИИ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ»

Основываясь на разработанной авторской методике исследования контекстного знания [12, 13], было проведено исследование формирования терминологического ландшафта междисциплинарного направления научных исследований «Информатизации научной деятельности». Для этого в качестве инструментов исследования были выбраны информационные системы Voyant-tools (<https://voyant-tools.org>) и Sketch Engine (<https://www.sketchengine.eu>), предназначенные для обработки текстов на естественном языке. В качестве опорных терминов были использованы слова «анализ», «система», «визуализация», «моделирование», «технологии», «исследование», «наука», «программирование». Эти термины входят в терминологическое ядро исследуемого направления, сформированное на предыдущем этапе. Кроме этого, в исследовательских целях были выбраны имена прилагательные

«виртуальный», «электронный», «цифровой». Они определяют смысловые оттенки используемых терминов. К тому же отличительной особенностью Sketch Engine является возможность в качестве исследуемых терминов использовать не только имена существительные.

Использование инструмента «Collocation» в Voyant-tools на корпусе текстов материалов конференций «Интернет и современное общество» (с 2011 по 2019 годы) и «Научный сервис в сети Интернет» (с 2015 по 2019 годы) позволило получить следующие результаты (таб. 1):

Таб. 1. Коллокации с представителями выявленного ядра направления научных исследований «Информатизации научной деятельности» (получены с использованием Voyant-tools)

ИССЛЕДУЕМЫЙ ТЕРМИН	КОЛЛОКАЦИИ (ПО УБЫВАНИЮ ЧАСТОТЫ ИСПОЛЬЗОВАНИЯ)
анализ	данных, текстов, результатов, информации, программ, документов, методов исследования
визуализация	данных, геоданных, результатов, документов, информации, коллекций, структуры, процессов
системы	науки, данных, мониторинга, научных, библиотеки, модели, технологии
моделирование	процессов, данных, систем, текстов, технологии
технологии	науки, информационные, электронные, исследований

С одной стороны, анализ полученных данных подтверждает соотнесение выявленных ранее терминов к терминологическому ядру (например: «анализ данных», «визуализация данных», «информационные технологии»). С другой – позволил включить в терминологический ландшафт новые термины (например: «анализ текстов», «анализ программ», «визуализация результатов», «визуализация геоданных», «визуализация процессов», «системы мониторинга», «системы – библиотеки», «моделирование процессов», «моделирование систем»). Пересечение использованных терминов со словами «наука», «технологии», «исследование» подтверждает, что выявленные термины, как и термины ядра, используются в контексте научных исследований и применения в них информационно-коммуникационных технологий.

Для исследования контекстов использования терминов «ресурс», «моделирование», «визуализация», «электронный», «программирование», «анализ» был использован инструмент Word Sketch, позволяющий выявлять коллокации и комбинации со словами, представляющими либо другие части речи, либо являющиеся модификаторами. В целом, были выявлены устойчивые сочетания, которые либо уже вписываются в терминологический ландшафт (например, являются представителями ядра), либо дополняют его новыми контекстами. Наглядно результаты представлены на

рис. 2-7.



Рис. 2. Коллокации термина «ресурс» (визуализация в Sketch Engine)



Рис. 3. Коллокации термина «моделирование» (визуализация в Sketch Engine)



Рис. 4. Коллокация термина «визуализация» (визуализация в Sketch Engine)



Рис. 5. Коллокации прилагательного «электронный» (визуализация в Sketch Engine)



Рис. 6. Коллокации термина «программирование» (визуализация в Sketch Engine)



Рис. 7. Коллокации термина «анализ» (визуализация в Sketch Engine)

Анализ полученных данных позволяет выявить следующие цепочки устойчивых сочетаний терминов, которые более чётко описывают смысловые оттенки основных терминов:

- ресурс – информационный, электронный, цифровой, научный;
- моделирование – математическое, компьютерное, тематическое, имитационное;
- визуализация – научная, компьютерная, мультимедийная, интерактивная;
- программирование – параллельное, системное;
- анализ – данные, текст, документ, результат, программа, корпус, сравнительный, семантический, интеллектуальный, автоматический, системный, кластерный.

Помимо этого, в отношении прилагательных выявлены основные термины, которые они характеризуют, например, «электронный» – правительство, ресурс, среда, участие, библиотека, демократия. Анализ цепочек подтверждает междисциплинарность направления научных исследований «Информатизации научной деятельности». Так, например, термин «анализ» связан с такими областями научных исследований, как лингвистика и языкознание, а не только с математическими дисциплинами или программированием.

VII. КОНТЕКСТНОЕ ЗНАНИЕ НЕТЕКСТОВОЙ МОДАЛЬНОСТИ В НАУЧНЫХ ИССЛЕДОВАНИЯХ

В рамках исследования терминологического ландшафта направления научных исследований «Информатизации научной деятельности» приходится сталкиваться не только с контекстами текстовой модальности, для которых была разработана соответствующая типология [13]. Контексты нетекстовой модальности, с одной стороны, являются объектами научных исследований, а, с другой, - подходы к их изучению образуют методологическую базу самих исследований. Поэтому для дальнейшего развития методов контекстного анализа в научных исследованиях представляется важным как изучение контекстов нетекстовой модальности, так и развитие методов и технологий их обработки и анализа.

Мы предлагаем анализировать текст и нетекстовые модальности информации сначала на более высоком (и более абстрактном) уровне – уровне обобщенных структурных инвариантов контекстуального знания, а затем редуцировать и специфицировать их применительно к конкретным дисциплинарным обстоятельствам и междисциплинарным отношениям (многоуровневый структурный анализ).

Следует отметить, что исследователи различают мономодальность и «более сложный и интегрированный вариант подачи информации» [22] – мультимодальность данных, документа, сообщения. Анализ контекстов корпуса мультимодальной информации, как правило, осуществляется по каждой из модальностей отдельно и часто независимо, поэтому для каждой из модальностей может быть предложена своя типология контекстов. В разработанной авторами типологии контекстов текстовой модальности выделяются группы контекстов, которые отличаются назначением, типом или характеристикой контекста и распадаются на отдельные виды. Использование тех или иных видов контекстов определяется задачами исследования и выбором технологий, а также используемым инструментарием для интеллектуального анализа данных.

Контексты нетекстовой модальности более сложная для изучения обобщенная группа, разнообразие которой полностью определяется свойствами, присущими той или иной модальности. Следует учитывать, что модальность задается принадлежностью данных к некоторому источнику, определяющему структуру, формат, структурно-функциональными связи данных, процедуры обработки и анализа данных [22]. Кресс Г. выделяет следующие типы модальности: обонятельная, осязательная, вкусовая, зрительная и слуховая и, соответственно, считает мультимодальными явлениями взаимодействия между вербальными текстами и изображениями, видео, речью и жестами, размером и цветом текста [33]. Контексты нетекстовой модальности могут быть рассмотрены и анализироваться сами по себе или как часть документа, содержащего одновременно данные/контексты нескольких модальностей. В этом случае идет речь о мультимодальных и мультязычных, как подмножестве, документах, системах, поиске, моделях. Таким образом, мультимодальным является любой текстовый документ,

объединяющий в себе «две или более семиотических систем для создания смысла» [35], «различные семиотические коды, требующие актуализации сразу нескольких перцептивных каналов, визуального и аудиального. ..., которые задействуются для составления и передачи сообщения» [22]. Технически, мультимодальность – это концепция интеграции информации из нескольких модальностей с целью прогнозирования. Мультимодальный поиск – это поиск, который позволяет осуществлять запросы/включать в запрос токены разной модальности последовательно или параллельно/одновременно (гибридный запрос).

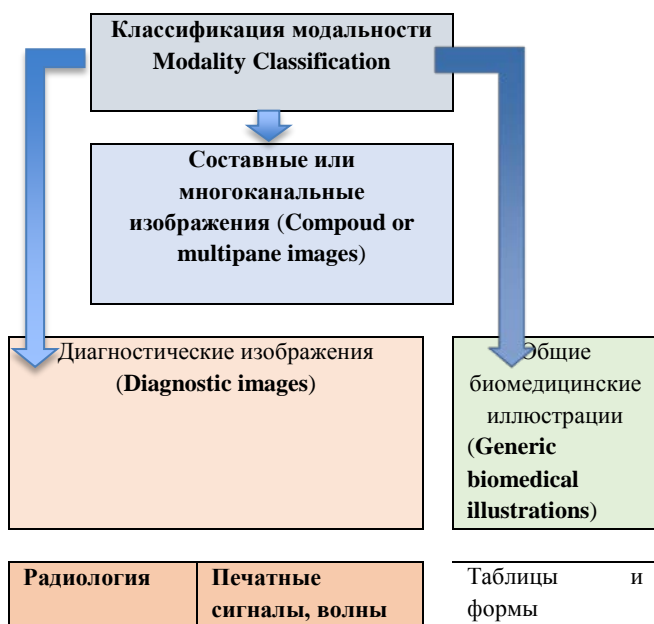
В качестве ориентира для восприятия сложности и разнообразия контекстов и как основа для развития типологии контекстов приводим токены нетекстовых модальностей применительно к модальностям базового, низкого уровня [9, 21, 22, 30, 31, 32, 33, 35, 36, 37]:

- a) метаданные документа – авторы, время, источник, рубрики, и т.д., метаданные файла (в зависимости от ситуации и исследовательского подхода могут рассматриваться как токены текстовой или нетекстовой модальности);
- b) токены, находящиеся внутри текста – ссылки, теги, словосочетания, именованные сущности, ссылки на изображениях, записи о действиях пользователей, и т.д.;
- c) языки: многоязычный набор текста; в системах кроссязычного и мультязычного тематического поиска запрос даётся на одном языке, а ответ может быть получен на других языках;
- d) типографика (художественная способ представления текста);
- e) рисунки, фотографии всех форматов;
- f) языковые конструкции и стили, которые включают использование символических языковых маркеров, таких как пунктуация, эмодзи (или другие пиктографические изображения), микротекст, мемы (вирусное изображение, видео или словесное выражение для мимикрии или юмористических целей), анимированные GIF-файлы (формат обмена графическими данными, который объединяет несколько изображений или «кадров» в один файл для передачи движения) [35];
- g) инфографика (infographics) (текст, встроенный вместе с изображением) – схемы, графики и т.п.
- h) схемы изображений (Image schemas) – скелетные концептуальные структуры, возникающие в результате восприятия, движений тела, манипулирования объектами, процессами и силами, такие как вверх / вниз, сила, противодействие или вход / выход. Схемы изображений в основном используются для объяснения связи между воплощенным опытом (особенно базовым пространственным опытом) и высшим познанием. Ощутимые различия схемы от рисунков состоят в том, как схемы формируют опыт и в том, что передаваемое ими знание зависит от дополнительного концептуального элемента – точки зрения. Добавление точки зрения как характеристики схемы изображения означает, что один и тот же тип структуры может давать разные

экспериментальные результаты. [32]. Таким образом, схемы отражают не только объекты, распознаваемые традиционными методами, но и мнения, представления с позиции определенного ракурса, а значит содержат дополнительный контекст, выявление и распознавание которого требует неординарных подходов;

- i) аудио и видео информация;
- j) интонационные инструменты воздействия, жесты, мимика, паузы, смех и т.д. образуют знаковую систему, дополняют или заменяют средства вербальной коммуникации. «Внешность оратора» [21], актора также можно рассматривать как токен и контекст.
- k) субтитры, реплики, звуковые эффекты, речь, акустические события и т.п. из аудио и визуальных каналов [30];
- l) движения камеры, методы съемки и тип действий считаются одними из наиболее важных визуальных характеристик видеоконтента, с помощью которых отличают различные жанры и техники съемок;
- m) цвета – для моделирования цвета и освещения могут извлекаться следующие визуальные признаки: гистограммы RGB, гистограмма значений: Соотношение RGB: Гистограмма насыщенности;
- n) лица и способ их изображения: характеристики, связанные с лицами: количество обнаруженных лиц в кадре и отношение площади ограничивающей рамки лица к общему размеру кадра.

Таким образом экспликация и понимание контекста, особенно контекста нетекстовой модальности, — это один из самых сложных аспектов модерации контента. Ситуация усугубляется особенностями распознавания и классификации модальности, налагаемые предметной областью и междисциплинарными направлениями исследований. В качестве примера предметной области, имеющей несомненную специфику, можно привести сферу здравоохранения и классификацию нетекстовой модальности базового уровня – изображений, применяемых в медицине (рис. 8).



Ультразвук	Электроэнцефалография	Список программ
Магнитный резонанс	Электрокардиография	Статистические цифры, графики, диаграммы
Компьютерная томография	Электромиография	Скриншоты
Рентген, 2D-рентгенография	Микроскопия	Блок-схемы
Ангиография	Световая микроскопия	Обзор систем
Позитронно-эмиссионная томография	Электронная микроскопия	Последовательность гена
Комбинированные методы в одном изображении	Просвечивающая микроскопия	Хроматография, Гель
Фотография в видимом свете	Флуоресцентная микроскопия	Химическая структура
Эндоскопия	3D реконструкции	Математика, формулы
Дерматология		Неклинические фотографии
Другие органы		Рисованные эскизы

Рис. 8. Иерархическая классификация модальностей медицинских изображений [33]

Создаваемые большие коллекции медицинских изображений играют важную роль прежде всего в медицинских исследованиях и принятии клинических решений. Одна из основных проблем заключается в том, что размер коллекций постоянно растет из-за увеличения доступности оборудования для визуализации в больницах. Это создает огромные хранилища ценной информации, которую во многих случаях трудно обрабатывать и управлять ею должным образом. Это является основанием для разработки инструментов для эффективного и действенного доступа к этому типу информации, объединения методов визуализации и взаимодействия с больничными и ведомственными информационными системами для управления хранением и распространением изображений среди медицинского персонала, исследователей, клиник и центров визуализации. Задача индексации и каталогизации этих коллекций традиционно выполнялась вручную. Это дорогостоящая и трудоемкая процедура, к тому же она подвержена ошибкам, связанным с человеческим фактором. Следовательно, существует острая потребность в автоматизированном индексировании коллекций медицинских изображений, чтобы улучшить возможность поиска и извлечения соответствующих изображений [33].

Системы поиска медицинских изображений традиционно основываются на тексте, относящегося к аннотации или подписям, связанных с изображением. Сочетание методов текстового и визуального поиска

улучшает производительность по сравнению с их отдельным использованием. Поэтому в последние годы получили развитие методы мультимодального поиска, где в качестве входных данных выступают запросы, включающие последовательно или параллельно токены разной модальности (гибридный запрос). В этом случае запросы состоят из текстовой части (т.е. текстового подзапроса) и / или образцов изображений (т.е. визуального подзапроса). Например, запросы могут содержать информацию о демографических характеристиках пациентов, ограниченном наборе симптомов и результатах медицинских обследований, включая визуализационные исследования. Базы данных медицинских изображений, используемые для поиска или в учебных целях, часто содержат изображения, полученные различными методами (например, рентген, компьютерная томография, ультразвук и т.д.), а их аннотации разнородны и не систематизированы. Это особенно верно для изображений, которые можно найти в различных онлайн-ресурсах, включая те, которые обращаются к онлайн-контенту журналов.

Модальность изображения – это фундаментальная визуальная характеристика изображения, которую можно использовать для повышения эффективности поиска. Однако аннотации или подписи, связанные с изображениями, часто не содержат информации о модальности и способе получения этого изображения [33]. Для адекватного представления изображения необходимо уметь распознавать признаки, которые позволяют уловить различные аспекты изображения (например, текстуру, формы, распределение цвета ...). Характеристики локального изображения имеют основополагающее значение для интерпретации изображения: в то время как глобальные элементы сохраняют информацию обо всем изображении, локальные элементы фиксируют детали. Таким образом, они более разборчивы в отношении проблемы межклассовой и внутриклассовой изменчивости. Цель состоит в том, чтобы правильно классифицировать модальность изображений, используя визуальную информацию из изображений и текста статьи, в которой встречается это изображение.

Подход [33] к классификации медицинских модальностей использует различные визуальные особенности в сочетании с текстовыми функциями, извлеченными из окружающего текстового содержания изображений. Также Дрокин И.С. и др. считают, что обработка данных доменно-специфична и зависит от типа данных (модальности и контексты) и источника данных. В качестве источника знаний помимо онтологий, таких как открытые биомедицинские онтологии (ОБО, англ. Open Biomedical Ontologies), или словарей может использоваться любой источник медицинских данных, из которого можно сформировать иерархию знаний, например медицинские схемы-гайдлайны и т.п. [9]

Модальность данных определяется на основе следующих соображений: выделяются текстовая и нетекстовая модальности, последняя может быть декомпозирована сначала на базовые, «низкие» уровни, а затем по каждой из модальности на уровни высшего

порядка (контексты), общие и задаваемые направлением исследований и задачами, стоящими перед исследователями.

В целом, анализ рассмотренных подходов к описанию контекстов нетекстовой модальности позволяет разбить их на следующие основные группы (таб. 2):

Таб. 2. Обобщённые группы контекстов нетекстовой модальности

ОБОБЩЕННАЯ ГРУППА: КОНТЕКСТЫ НЕТЕКСТОВОЙ МОДАЛЬНОСТИ	
Группы	Тип или характеристика контекста
Изображения	рисунки, фотографии, инфографика (infographics), схемы изображений (Image schemas), типографика
Аудио контент	звуки и сигналы
Видео контент	визуальные характеристики видеоконтента
Другие органы чувств	обонятельная, осязательная, вкусовая
Языки	многоязычный набор текста, языковые конструкции и стили,
Другие виды контента	метаданные документа и файла, токены, находящиеся внутри текста встроенные в текст

VIII. Выводы

Исходя из полученных результатов проведённого исследования можно сделать следующие основные выводы:

- отобранные источники данных для проведения исследования релевантны исследуемому направлению исследований «Информатизация научной деятельности»;
- подтверждена междисциплинарность исследуемого направления «Информатизация научной деятельности»;
- выявлены основные научные области и направления исследований, входящие в направление междисциплинарных исследований «Информатизация научной деятельности»;
- сформировано ядро терминологической базы направления междисциплинарных исследований «Информатизация научной деятельности»;
- отобраны термин-концепты, которые после дополнительных исследований могут быть включены в терминологическую базу направления междисциплинарных исследований «Информатизация научной деятельности»;
- выявлены синонимичные термин-концепты, которые будут включены в разрабатываемый в рамках исследования тезаурус терминологической базы направления междисциплинарных исследований «Информатизация научной деятельности»;
- исследован терминологический ландшафт исследуемого направления;
- уточнён контекст и смысловые оттенки некоторых терминов ядра исследуемого направления;
- поставлена проблема необходимости учёта и исследования мультимодального контекстного знания, являющегося предметом научных исследований.

Практическое значение результатов проведённого исследования определяется тем, что используемые в нём подходы и методы войдут в разрабатываемый авторами учебно-методический комплекс (УМК) «Технологии

извлечения и интеллектуального анализа данных в научных исследованиях», направленный на формирование компетенций магистрантов, соответствующих образовательному стандарту направления 09.04.03 «Прикладная информатика» Университета ИТМО, общей характеристике образовательной программы «Цифровые технологии умного города», созданной на базе профессиональных стандартов 06.022 «Системный аналитик» и 06.015 «Специалист по информационным системам». Внедрение разрабатываемого УМК позволит повысить аналитические и инструментальные компетенции преподавателей и сотрудников, выполняющих научное руководство магистрантами и аспирантами. УМК также может быть предложен для использования в научно-исследовательской работе магистрантов и аспирантов программ магистратуры Университета ИТМО, а также и других вузов.

Дальнейшее исследование предполагает выявление тенденций развития рассматриваемого направления научных исследований на основе анализа динамики публикационной активности и построения соответствующих трендов.

Выявленные в ходе исследования термин-концепты имеют англоязычные аналоги, которые будут использованы в качестве опорных термин-концептов при поиске, экспликации и анализе мирового англоязычного научного дискурса по исследуемому направлению.

Сформированный в результате исследования тезаурус направления междисциплинарных исследований «Информатизация научной деятельности» будет структурированно описан в формате метаданных спецификации Dublin Core и представлен в машиночитаемой форме для обеспечения свободного доступа к нему представителями научного сообщества.

Анализ подходов к исследованию и применению мультимодальных контекстов ставит задачу дальнейшего их изучения. А формирование и анализ категориального аппарата междисциплинарных направлений исследований требует дальнейшего уточнения в части типологии контекстов нетекстовой модальности.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект 18-011-00923-а) и Благотворительного фонда Владимира Потанина (проект ГК200000654).

БИБЛИОГРАФИЯ

- [1] Баканов В.М. Использование системы добровольных распределенных вычислений для оптимизации ярусно-параллельной формы информационных графов алгоритмов // Современное образование: содержание, технологии, качество. 2015. № 2. С. 129-130.
- [2] Бородин Л.И. Историческая информатика сегодня: вызовы "цифровой эпохи" // Информационный бюллетень ассоциации история и компьютер. 2014. №42. С. 3-6.
- [3] Володин А. Ю. «Цифровая история»: ремесло историка в цифровую эпоху // История. 2015. Т.6. Выпуск 8 (41). DOI: 10.18254/S0001228-9-1.
- [4] Володин А.Ю. Цифровые гуманитарные науки (Digital Humanities): вызовы и тупики междисциплинарности // «Стены и мосты»–IV: междисциплинарные исследования в истории. М.: Академический проект, 2016. С. 139-147.
- [5] Гарскова И.М. Историческая информатика: сравнительный анализ национальных моделей // Информационный бюллетень ассоциации история и компьютер. 2018. №47. С. 25-26.
- [6] Гартман Т. Н., Советин Ф. С. Аналитический обзор современных пакетов моделирующих программ для компьютерного моделирования химико-технологических систем // Успехи в химии и химической технологии. 2012. №11 (140). С. 117-120.
- [7] Глинский Б.М., Кучин Н. В., Черных И. Г., Орлов Ю. Л., Подколотный Н. Л., Лихошвай В. А., Колчанов Н. А. Суперкомпьютерные технологии в решении задач биоинформатики // Программные системы: теория и приложения. 2015. №4 (27). С. 99-122. DOI: 10.25209/2079-3316-2015-6-4-99-112.
- [8] Долгов В.И., Неласая А.В. Методы увеличения скорости криптографических преобразований на эллиптических кривых // Радиотехника. Информатика. Управление. 2004. № 2. С. 72-78.
- [9] Дрокин И.С., Бухвалов О.Л., Сорокин С.Ю. Способ формирования математических моделей пациента с использованием технологий искусственного интеллекта // Патент RU 2 720 363 С2. 29.12.2017.
- [10] Ежова Н.А., Соколинский Л.Б. Обзор моделей параллельных вычислений // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2019. Том 8. №3. С. 58-91. DOI: 10.14529/cmse190304.
- [11] Желтов С.А. Адаптация метода Шермана–Лемана решения задачи факторизации к вычислительной архитектуре CUDA // История и архивы. 2012. №14 (94). С. 84-91.
- [12] Кононова О.В., Ляпин С.Х., Прокудин Д.Е. Исследование терминологической базы междисциплинарного научного направления «цифровая экономика» с использованием инструментов контекстного анализа // International Journal of Open Information Technologies. 2018. Том 6, № 12. С. 57-66.
- [13] Кононова О.В., Прокудин Д.Е. Подход к извлечению, экспликации и представлению контекстного знания при изучении развивающихся междисциплинарных направлений исследований // International Journal of Open Information Technologies. 2020. Том 8, № 1. С. 90-101. URL: <http://injoit.org/index.php/ji/article/view/882/844>.
- [14] Кононова О.В., Прокудин Д.Е., Смирнова П.В. Технологии изучения контекстного знания при исследованиях основных направлений геймификации в городском развитии // Информационное общество: образование, наука, культура и технологии будущего. Выпуск 3 (Труды XXII Международной объединенной научной конференции «Интернет и современное общество», IMS-2019, Санкт-Петербург, 19–22 июня 2019 г. Сборник научных трудов). СПб: Университет ИТМО, 2019. С. 53–66. DOI: 10.17586/2587-8557-2019-3-53-66.
- [15] Концепция информатизации Казанского государственного университета. <http://old.kpfu.ru/uit/index.php?id=4&idm=0&num=4>
- [16] Кручинин С. В. Математическое и компьютерное моделирование в политологии и политике (обзор) // Научно-исследовательские публикации. 2017. №4 (42). С. 34-41.
- [17] Можаяева Г.В. Digital Humanities: цифровой поворот в гуманитарных науках // Гуманитарная информатика. 2015. № 9. С. 8-23. DOI: 10.17223/23046082/9/1_15
- [18] Омаров М. Д. Аналитический обзор методологии компьютерного моделирования // Вестник ДГТУ. Технические науки. 2015. Том 36. №1. С. 84-89. DOI: 10.21822/2073-6185-2015-36-1-84-89.
- [19] Пеньков С.В. Историческая информатика: история и современность // Научный альманах. 2019. №12-2 (62). С. 71-73.
- [20] Семичевская Н.П., Соловцова Л.А., Питулина П.И. Применение параллельных вычислений в современных методах криптоанализа // Ученые заметки ТОГУ. 2016. Т. 7. № 4-1. С. 142-148. URL: http://pnu.edu.ru/media/ejournal/articles-2016/TGU_7_194.pdf.
- [21] Смелик Н. Д., Фильченков А. А. Мультимодальная тематическая модель текстов и изображений на основе использования их векторного представления // Машинное обучение и анализ данных, 2016. Том 2. No 4. С. 421-441. DOI: 10.21469/22233792.2.4.05
- [22] Сорокина Ю. В. Понятие мультимодальности и вопросы анализа мультимодального лекционного дискурса // Филологические науки. Вопросы теории и практики. 2017. № 10. Ч. 1. С. 168-170.
- [23] Спицина А. М., Орлов Ю.Л., Подколотная Н.Н., Свичкарев А.В., Дергилев А.И., Чен Минг, Кучин Н.В., Черных И.Г., Глинский

- Б.М. Суперкомпьютерный анализ геномных и транскриптомных данных, полученных с помощью технологий высокопроизводительного секвенирования ДНК // Программные системы: теория и приложения. 2015. №1 (24). С. 157-174. DOI: 10.25209/2079-3316-2015-6-1-157-174.
- [24] Терехов А.Н., Сепман В.Ю., Кияев В.И., Комаров С.Н. Концепция информатизации Санкт-Петербургского государственного университета. URL: https://www.math.spbu.ru/user/ant/all_articles/068_Terekhov_Kiyaev_Komarov_Koncept_Informat.pdf.
- [25] Цифровая гуманитаристика: ресурсы, методы, исследования: материалы Междунар. науч. конф. (г. Пермь, 16–18 мая 2017 г.): в 2 ч. / Перм. гос. нац. исслед. ун-т. Пермь, 2017. Ч. 1. 175 с.
- [26] Цифровая гуманитаристика: ресурсы, методы, исследования: материалы Междунар. науч. конф. (г. Пермь, 16–18 мая 2017 г.): в 2 ч. / Перм. гос. нац. исслед. ун-т. Пермь, 2017. Ч. 2. 208 с.
- [27] Шамакина А.В. Обзор технологий распределенных вычислений // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2014. Том 3. №3. С. 58-85. DOI: 10.14529/cmse140304.
- [28] Якимец В.Н., Курочкин И.И. Развитие проектов добровольных распределенных вычислений на основе дорожных карт и многопараметрических оценок // International Journal of Open Information Technologies. 2020. Т. 8. №1. С. 1-8. URL: <http://injoit.org/index.php/j1/article/view/864/831>.
- [29] Ямщиков О. Н. Компьютерное моделирование в травматологии и ортопедии (обзор литературы) // Вестник российских университетов. Математика. 2014. №6. С. 1974-1979. URL: <http://journals.tsutmb.ru/go/1810-0198/2014/6/1974-1979>.
- [30] Bougiatiotis K., Giannakopoulos T. Enhanced movie content similarity based on textual, auditory and visual information // Expert Systems with Applications. 2018. Volume 96. P. 86-102. DOI: 10.1016/j.eswa.2017.11.050
- [31] Cromley J. G., Kunze A. J., Parpuce Dane A. Multi-text multi-modal reading processes and comprehension // Learning and Instruction. 2021. Volume 71. 101413. DOI: 10.1016/j.learninstruc.2020.101413
- [32] Dancygier B., Vandelanotte L. Image-schematic scaffolding in textual and visual artefacts // Journal of Pragmatics. 2017. Volume 122. P. 91-106. DOI: 10.1016/j.pragma.2017.07.013
- [33] Dimitrovski I., Koccev D., Kitanovski I., Loskovska S., Džeroski S. Improved medical image modality classification using a combination of visual and textual features // Computerized Medical Imaging and Graphics. 2015. Volume 39. P. 14-26. DOI: 10.1016/j.compmedimag.2014.06.005
- [34] Kumar A., Srinivasan K., Cheng W.-H., Zomaya A.Y. Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data // Information Processing & Management. 2020. Volume 57. Issue 1. 102141. DOI: 10.1016/j.ipm.2019.102141
- [35] Wang K., Meng W., Li S., Yang S. Multi-Modal Mention Topic Model for mentionee recommendation // Neurocomputing. 2019. Volume 325. P. 190-199. DOI: 10.1016/j.neucom.2018.10.024
- [36] Xu J., Huang F., Zhang X., Wang S., Li C., Li Z., He Y. Visual-textual sentiment classification with bi-directional multi-level attention networks // Knowledge-Based Systems. 2019. Volume 178. P. 61-73. DOI: 10.1016/j.knsys.2019.04.018

Кононова Ольга Витальевна, канд. экон. наук, доцент, Университет ИТМО, ORCID 0000-0001-6293-7243 (e-mail: kononolg@yandex.ru).

Прокудин Дмитрий Евгеньевич, докт. филос. наук, доцент, Санкт-Петербургский государственный университет; Университет ИТМО, ORCID 0000-0002-9464-8371 (e-mail: hogben.young@gmail.com).

Technologies for extraction, explication and analysis of multimodal contextual knowledge in the study of the terminological base of the interdisciplinary research direction "Informatization of scientific activity"

O.V. Kononova, D.E. Prokudin

Abstract— Digital transformations, which are the basis for the development of modern society, often outstrip the processes of transformation of science. At the same time, they force science to constantly improve both technologies and software tools used in research. In addition, the development of Informatization and digitalization processes leads to the formation of new interdisciplinary scientific directions that reflect the dynamics of social development. Today, there are already established areas of use of information and communication technologies in scientific activities. However, "Informatization of scientific activity" as an integral direction of scientific research has not yet been formed. Therefore, the research is aimed at identifying the main thematic areas that make up the direction of interdisciplinary research "Informatization of scientific activity", determining the core of its terminology base. Based on the application of the approach developed by the authors (synthetic method) to research on the development of topics and the conceptual and terminological landscape of interdisciplinary scientific areas, the main term concepts involved in the formation of the thesaurus are identified. The terminological landscape of the research area is also clarified by identifying the contexts and semantic shades of the terms of the core of the research area.

The study established the multimodality of contextual knowledge used in scientific research. A primary analysis of the representation of non-text modality contexts is performed.

The study was based on data obtained from various digital sources that represent scientific publications over the past 10 years. Voyant-Tools and Sketch Engine systems are used as tools for explication and analysis of contextual knowledge.

Keywords— contextual knowledge, contextual search, multimodality, scientific activity, Informatization, interdisciplinary research areas, terminology base, thesaurus, term-concept, Voyant-Tools, Sketch Engine.

REFERENCES

- [1] Bakanov V.M. Ispol'zovanie sistemy dobrovol'nyh raspredelennyh vychislenij dlja optimizacii jarusno-parallel'noj formy informacionnyh grafov algoritmov // *Sovremennoe obrazovanie: sodержanie, tehnologii, kachestvo*. 2015. № 2. P. 129-130.
- [2] Borodkin L.I. Istoricheskaja informatika segodnja: vyzovy "cifrovoy jepohi" // *Informacionnyj bjulleten' asociacii istorija i komp'juter*. 2014. №42. P. 3-6.
- [3] Volodin A. Digital History: the Craft of Historian in the Digital Age // *Istoriya*. 2015. Vol. 6. Iss. 8 (41). DOI: 10.18254/S0001228-9-1.
- [4] Volodin A.Ju. Cifrovye gumanitarnye nauki (Digital Humanities): vyzovy i tupiki mezhdisciplinarnosti // «Steny i mosty»—IV: mezhdisciplinarnye issledovanija v istorii. M.: Akademicheskij projekt, 2016. P. 139-147.
- [5] Garskova I.M. Istoricheskaja informatika: sravnitel'nyj analiz nacional'nyh modelej // *Informacionnyj bjulleten' asociacii istorija i komp'juter*. 2018. №47. P. 25-26.
- [6] Gartman T. N., Sovetin F. S. Analiticheskij obzor sovremennyh paketov modelirujushhijh programm dlja komp'juternogo modelirovanija himiko-tehnologicheskijh sistem // *Uspehi v himii i himicheskij tehnologii*. 2012. №11 (140). P. 117-120.
- [7] Glinskiy B., Kuchin N., Chernykh I., Orlov Yu., Podkolodnyi N., Likhoshvai V., Kolchanov N. Bioinformatics and High Performance Computing // *Program systems: theory and applications*. 2015. №4 (27). P. 99-122. DOI: 10.25209/2079-3316-2015-6-4-99-112.
- [8] Dolgov V.I., Nelasaja A.V. Metody uvelichenija skorosti kriptograficheskijh preobrazovanij na jellipticheskijh krivyh // *Radioelektronika. Informatika. Upravlenie*. 2004. № 2. P. 72-78.
- [9] Drokin I.S., Bukhvalov O.L., Sorokin S.Yu. Sposob formirovaniya matematicheskijh modelej patsienta s ispol'zovanijem tehnologii iskusstvennogo intellekta // *Patent RU 2 720 363 C2*. 29.12.2017
- [10] Ezhova N. A., Sokolinsky L. B. Survey of parallel computation models // *Computational Mathematics and Software Engineering*. — 2019. Vol. 8. №3. P. 58-91. DOI: 10.14529/cmse190304.
- [11] Zheltov S. Adaptation factorization problem solution by Sherman–Lehman method to the computing architecture CUDA // *History and Archives*. 2012. №14 (94). P. 84-91.
- [12] Kononova O. V., Lyapin S. Kh., Prokudin D. E. Studying the Interdisciplinary Terminological Landscape of Digital Economy with the Use of Contextual Analysis Tools // *International Journal of Open Information Technologies*. 2018. Vol. 6. № 12. P. 57-66.
- [13] Kononova O. V., Prokudin D. E. An approach to extraction, explication and presentation of contextual knowledge in the study of developing interdisciplinary research areas // *International Journal of Open Information Technologies*. 2020. Vol. 8, № 1. P. 90-101. URL: <http://injoit.org/index.php/j1/article/view/882/844>.
- [14] Kononova O. V., Prokudin D. E., Smirnova P.V. Approach to Use of Network Scientific Environment for Studying the Interdisciplinary Terminological Landscape of Digital Economy // *Information Society: Education, Science, Culture and Technology of Future*. Issue 3. P. 53–66. DOI: 10.17586/2587-8557-2019-3-53-66
- [15] Konceptcija informatizacii Kazanskogo gosudarstvennogo universiteta. — <http://old.kpfu.ru/uit/index.php?id=4&idm=0&num=4>
- [16] Kruchinin S.V. Mathematical and computer modeling in Political Science and Politics (review) // *JSPR*. 2017. №4 (42). P. 34-41.
- [17] Mozhaeva G.V. Digital Humanities: digital turn in the humanities // *Humanitarian Informatics*. 2015. № 9. C. 8-23. DOI: 10.17223/23046082/9/1.
- [18] Omarov M.D. Analytical review of the methodology of computer modeling // *Herald of Dagestan State Technical University. Technical Sciences*. 2015. Vol. 36. №1. P. 84-89. DOI: 10.21822/2073-6185-2015-36-1-84-89.
- [19] Penkov S.V. Historical information science: history and modernity // *Science Almanac*. 2019. №12-2 (62). P. 71-73.
- [20] Pitulina P. I., Semichevskaya N. P., Solovtsova L. A. Application of parallel computing in the modern method of cryptanalysis // *Scientists notes PNU*. 2016. Vol. 7. № 4-1. P. 142-148. URL: http://pnu.edu.ru/media/ejournal/articles-2016/TGU_7_194.pdf.
- [21] Smelik N. D., Filchenkov A. A. Multimodal topic model for texts and images utilizing their embeddings // *Machine Learning and Data Analysis*. 2016. Volume 2. Issue 4. C. 421-441. DOI: 10.21469/22233792.2.4.05
- [22] Sorokina Yu. V. Notion of multimodality and issues of multimodal lecture discourse analysis // *Philology. Theory & Practice*. 2017. № 10. Part 1. P. 168-170.
- [23] Spitsina A. M., Orlov Yu. L. et al. Supercomputer analysis of genomics and transcriptomics data revealed by high-throughput DNA

- sequencing // Program systems: theory and applications. — 2015. №1 (24). P. 157-174. DOI: 10.25209/2079-3316-2015-6-1-157-174.
- [24] Terehov A.N., Sepman V.Ju., Kijaev V.I., Komarov S.N. Konceptija informatizacii sankt-peterburgskogo gosudarstvennogo universiteta. https://www.math.spbu.ru/user/ant/all_articles/068_Terehov_Kiyaev_Komarov_Koncept_Informat.pdf.
- [25] Cifrovaja gumanitaristika: resursy, metody, issledovanija: materialy Mezhdunar. nauch. konf. (g. Perm, 16–18 may 2017 g.): v 2 ch. / Perm. gos. nac. issled. un-t. Perm, 2017. Ch. 1. 175 p.
- [26] Cifrovaja gumanitaristika: resursy, metody, issledovanija: materialy Mezhdunar. nauch. konf. (g. Perm, 16–18 may 2017 g.): v 2 ch. / Perm. gos. nac. issled. un-t. Perm, 2017. Ch. 2. 208 p.
- [27] Shamakina A. V. Survey on distributed computing technologies // Computational Mathematics and Software Engineering. 2014. Vol. 3. №3. P. 58-85. DOI: 10.14529/cmse140304.
- [28] Yakimets V. N., Kurochkin I. I. Development of voluntary distributed computing projects based on roadmaps and multi-parameter assessments // International Journal of Open Information Technologies. 2020. Vol. 8. № 1. P. 1-8. URL: <http://injoit.org/index.php/j1/article/view/864/831>.
- [29] Yamshchikov O.N. Computer modeling in traumatology and orthopedics (literature review) // Tambov University Reports. Series: Natural and Technical Sciences. 2014. №6. P. 1974-1979. URL: <http://journals.tsutmb.ru/go/1810-0198/2014/6/1974-1979>.
- [30] Bougiatiotis K., Giannakopoulos T. Enhanced movie content similarity based on textual, auditory and visual information // Expert Systems with Applications. 2018. Volume 96. P. 86-102. DOI: 10.1016/j.eswa.2017.11.050
- [31] Cromley J. G., Kunze A. J., Parpucu Dane A. Multi-text multi-modal reading processes and comprehension // Learning and Instruction. 2021. Volume 71. 101413. DOI: 10.1016/j.learninstruc.2020.101413
- [32] Dancygier B., Vandelandotte L. Image-schematic scaffolding in textual and visual artefacts // Journal of Pragmatics. 2017. Volume 122. P. 91-106. DOI: 10.1016/j.pragma.2017.07.013
- [33] Dimitrovski I., Kocev D., Kitanovski I., Loskovska S., Džeroski S. Improved medical image modality classification using a combination of visual and textual features // Computerized Medical Imaging and Graphics. 2015. Volume 39. P. 14-26, DOI: 10.1016/j.compmedimag.2014.06.005
- [34] Kumar A., Srinivasan K., Cheng W.-H., Zomaya A.Y. Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data // Information Processing & Management. 2020. Volume 57. Issue 1. 102141. DOI: 10.1016/j.ipm.2019.102141
- [35] Wang K., Meng W., Li S., Yang S. Multi-Modal Mention Topic Model for mentionee recommendation // Neurocomputing. 2019. Volume 325. P. 190-199. DOI: 10.1016/j.neucom.2018.10.024
- [36] Xu J., Huang F., Zhang X., Wang S., Li C., Li Z., He Y. Visual-textual sentiment classification with bi-directional multi-level attention networks // Knowledge-Based Systems. 2019. Volume 178. P. 61-73. DOI: 10.1016/j.knsys.2019.04.018