

Исследование терминологической базы междисциплинарного научного направления «цифровая экономика» с использованием инструментов контекстного анализа

О.В. Кононова, С.Х. Ляпин, Д.Е. Прокудин

Аннотация — Динамика общественного развития приводит к формированию различных междисциплинарных научных направлений, влияющих как на само общественное развитие, так и на трансформации самих научных исследований, ориентированных на это развитие. Развитие таких междисциплинарных научных направлений как, например, «цифровая экономика: электронное государственное управление и умные технологии»; «социально-этические аспекты цифровой экономики: геймификация в социальной и научно-образовательной сферах»; «информатизация научной деятельности» в цифровую эпоху связана с развитием технологий информационного общества. Также технологии являются одним из основных инструментов, позволяющим оперативно и качественно проводить исследования динамики этого развития. Наше исследование проводится на массивах информации, полученных из различных цифровых источников, отражающих как последние научные публикации, так и содержание общественно-политического дискурса (например, публикации в СМИ).

В данном исследовании показана возможность применения разрабатываемого синтетического метода для решения задач отбора цифровых информационных ресурсов, извлечения из них контекстного знания и на его основе уточнения терминологической базы формирующегося междисциплинарного научного направления «Цифровая экономика: электронное государственное управление и умные технологии». В качестве инструмента поиска, экспликации и анализа данных используется информационная среда для научных исследований «Humanitarianana», технологической основой которой является распределенная информационная система T-Libra.

Ключевые слова — контекстное знание, контекстный поиск, цифровая экономика, цифровые технологии, государственное управление, умные технологии, распределенная сетевая среда, информационные ресурсы,

междисциплинарные научные направления,
терминологическая база

I. ВВЕДЕНИЕ

В современных науковедческих исследованиях актуализируется задача анализа содержания перспективных междисциплинарных научных направлений, что позволяет прогнозировать востребованность результатов исследований этих направлений (знаний) в различных областях общественного бытия. В связи с этим одной из основных тенденций современных науковедческих исследований является выявление (экспликация) контекстного знания путем применения методов, подходов, технологий и инструментов Digital Humanities. К ним можно отнести следующие:

- методы и технологии извлечения контекстных знаний из больших массивов данных (Data Mining, Big Data);
 - методы и технологии исследования тематических и смысловых контекстов, автоматически извлекаемых из неструктурированных источников (контекстный поиск).
- Применение Digital Humanities в науковедческих исследованиях обусловлено объективными факторами, связанными с развитием современных междисциплинарных научных направлений исследований:
- стремительное увеличение объемов порождаемой в рамках этих направлений информации;
 - неформализованность и разнородность представления этой информации;
 - распределенность мест её хранения и доступа;
 - избыточность и многозначность используемой терминологической базы.

Терминологическая база междисциплинарных научных направлений формируется как через перенос терминов из одной предметной области в другую без необходимой интерпретации и адаптации, так и через непосредственное заимствование из иноязычных научных источников.

Привлечение в науковедческие исследования технологий и инструментов Digital Humanities обусловлено также широким развитием сетевого и распределенного доступа к информационным ресурсам (научным данным и знаниям) и высокой скоростью

Статья получена 22.10.2018

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

Ляпин Сергей Хамзеевич, Архангельский краеведческий музей, канд. филос. наук (e-mail: lyapins@yandex.ru).

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

обновляемости знаний. Наблюдается отставание развития наукометрических и науковедческих дисциплин от темпов роста терминологической и категорической базы междисциплинарных научных направлений, неуправляемо формируемой научными школами, группами, отдельными исследователями. Многозначность терминологии и неструктурированность информации даже при свободном к ней доступе не позволяет оперативно отслеживать возникающие тенденции и связи. Это приводит к потерям значимой части научных знаний и гипотез, не получившим сразу широкого распространения. Этой проблематикой и определяется актуальность решения проблемы своевременного изучения формирования категориально-понятийной базы развивающихся междисциплинарных научных направлений с привлечением методов и инструментов Digital Humanities.

II. ЦЕЛЬ И ЗАДАЧИ ИССЛЕДОВАНИЯ

Основной целью исследования является изучение возможностей экспликации научного знания, полученного из цифровых информационных ресурсов, в распределённой сетевой среде с использованием методов и технологий извлечения и анализа контекстного знания.

Для достижения поставленной цели должны быть решены следующие основные задачи:

- обоснование выбора прикладной среды для обеспечения целей исследования в соответствии с принятой моделью экспликации контекстного знания;
- формирование семантического ядра (ключевых понятий) предметной области междисциплинарного научного направления «Цифровая экономика: электронное государственное управление и умные технологии»;
- исследование возможностей использования цифровых информационных ресурсов для формирования массивов научных текстов, релевантных предметной области (Scopus, WoS, платформы издательств, EBSCO, JSTOR, Академия Google, РИНЦ, Киберленинка и пр.). Характеристика ресурсов, критерии, влияющие на выбор, политематичность и т.д.;
- оценка массивов текстов распределённой сетевой среды на релевантность заданной предметной области (тестирование по основным ключевым понятиям);
- разработка структурированного описания контекстного знания для изучения развития тезауруса междисциплинарного научной области «цифровая экономика»;
- разработка структуры и мета-описания тезауруса междисциплинарной научной области «цифровая экономика».

III. ПОДХОДЫ, МЕТОДЫ И ТЕХНОЛОГИИ ИССЛЕДОВАНИЯ

Существующие методы и соответствующие им инструментарию поиска, экспликации и анализа контекстного знания, реализованные в отдельных информационных системах и веб-сервисами остаются малоизвестными и зачастую невостребованными из-за отсутствия содержательной информации как о самом

инструменте, так и о методах и алгоритмах его применения в науке. Слабая востребованность инструментарию обусловлена отсутствием тематических разработок по отбору и систематизации мета-описаний, тезаурусов, онтологий предметных областей, являющихся «базой данных и знаний» этих инструментов, а также реализацией в каждом отдельном инструментарию ограниченного набора методов представления, поиска и интерпретации информации.

В связи с этим для извлечения и контекстного анализа русскоязычных текстов мы останавливаемся на доступной распределённой сетевой среде «Humanitarianana», в которой в качестве инструмента контекстного поиска используется информационная система T-Libra [15], [17], [18]. В информационной системе реализован инструментарию контекстного поиска в русскоязычных и англоязычных текстах, анализа горизонтальных и вертикальных макро- и микро-контекстов и построения терминограмм. «Humanitarianana» может функционировать в режиме локальной сети и в режиме распределённой информационной среды с возможностью обращения ко всем ресурсам с любого из серверов организаций-участниц. Это позволяет привлекать к исследованию необходимое число участников – как исследователей, так и исследовательских организаций. К ним, в первую очередь, относятся организации-владельцы крупных информационных массивов и тематических коллекций материалов в гуманитарной и социальной сферах (музеи, вузы, информационные центры, библиотеки). Информационная система в настоящее время активно развивается разработчиками и широко используется (Музеи Московского Кремля – научно-справочная библиотека; Архангельский краеведческий музей – научная библиотека; Университет ИТМО – информационно-ресурсный центр для извлечения знаний из гуманитарных текстов на основе использования распределённой сетевой среды «Humanitarianana»).

Разрабатываемый комплексный подход, названный нами синтетическим методом, выводит контент-анализ на иной уровень. Синтетический метод извлечения контекстного знания ориентирован на решение проблем выделения и экспликации научного контента с использованием систем продвинутого полнотекстового и мультимодального поиска в сетевых распределённых средах [14]. В традиционном контент-анализе первичными являются целевая функция и категории анализа, вторичными – получаемые обобщенно-текстовые единицы анализа. Для синтетического метода первичен обобщенный текст (с элементами мультимодальной информации), вторичен получаемый контент – структурированное описание контекстуального знания. Можно сказать, что традиционный контент-анализ и предлагаемый синтетический анализ контекстного знания являются дополняющими друг друга методами и технологиями изучения содержательных и смысловых информационных контекстов. Обобщенный текст (текст + мультимодальная информация) в этом случае является генератором эксплицируемых контекстов и структур контекстного знания. Инструментом генерации являются гибкие функциональные структуры мультимодальных запросов.

В отличие от уже существующих разработок ожидаемые результаты применения синтетического метода позволят глубже изучать структуры и функции знания по междисциплинарным научным исследованиям, использовать контекстные структуры для интерпретации научных текстов, разработки алгоритмов, компьютерных программ интеллектуального поиска и искусственного интеллекта [13].

IV. ТЕХНОЛОГИИ АВТОМАТИЗИРОВАННОГО ИЗВЛЕЧЕНИЯ И АНАЛИЗА КОНТЕКСТНОГО ЗНАНИЯ

Современные исследования базируются на использовании текстовых источников различной природы, в первую очередь цифровых научных ресурсов. Поэтому технологии экспликация контекстного знания требуют постоянного совершенствования. Поиск и анализ эксплицируемых смысловых и тематических контекстов важны для широкого круга задач в сферах науки и образования, государственного управления и бизнеса, востребованы как на мировом уровне, так и в России. Однако, контекстный поиск как технология до сих пор не получил широкого распространения т.к. применение этого вида поиска подразумевает наличие вполне структурированного набора источников и четкое понимание предмета поиска.

Контекстное знание (т.е. знание, содержащееся в различных контекстах, полученных в результате полнотекстовых запросов) может существовать в различных видах и представлено в разных формах.

Задачи изучения контекстного знания решаются через реализацию различных подходов:

1) технологического, который направлен на разработку информационных систем и реализации в поисковых системах алгоритмов контекстуального поиска [4], [16], [20].

2) семантического, направленного на разработку и применение лингвистических методов к анализу тестов и выявлению в них определённых смыслов [1], [9], [19], [22].

3) содержательного, состоящего в прикладном применении алгоритмов поиска и анализа информации в информационных системах с последующими количественной обработкой текстов и качественном анализе содержащихся в них смыслов из определённых предметных областей [3].

Исследование контекстов проводятся как в рамках

– традиционного контент-анализа, который сочетает в себе методы и технологии качественно-количественного анализа документов для выявления или измерения различных фактов и тенденций, отраженных этими документами. Результаты исследований на базе одного из трех подходов с применением традиционного контент-анализа отражены в ряде публикаций [3], [5], [7]-[10], [12], [16], [17], [19], [20];

– так и комплексного подхода, названного авторами синтетическим методом [2], [13], [14].

Синтетический метод предполагает применение расширенного набора инструментов и сервисов электронной библиотеки распределенного доступа T-Libra (<http://www.tlibra.ru>) для русскоязычных текстов и

системы с аналогичным функционалом Voyant-Tools (<https://voyant-tools.org/>), Tropes High Performance Text Analysis (<http://www.semantic-knowledge.com/>):

– автоматическая кластеризация результатов абзацно-ориентированного запроса с обратной связью с поисковым запросом (позволяет осуществлять кластеризацию запроса и управлять ее параметрами);

– многослойный тематический запрос с вариацией используемых слоев, позволяющий выделять аспекты эксплицируемой темы (от 2 до 8 аспектов).

– фокусировка полнотекстового запроса (позволяющая задавать расстояние между поисковыми терминами в искомом абзаце, и находить оптимальное соотношение между полнотой и точностью поиска);

– каскадный поиск (результаты одного запроса автоматически входят в поисковый образ другого запроса; позволяет осуществлять структурную модуляцию запроса для выявления новых элементов контекстного знания;

– гибридный квазисемантический поиск (одновременно по описаниям ресурса, взятым из каталога, и по полным текстам; используется для мультимодального поиска – например, по описаниям музейных артефактов и по полным текстам библиотеки);

– тезаурусный поиск (абзацный поиск с автоматическим включением в поисковый образ функциональной структуры тезауруса; используется для автоматического расширения культурного контекста в ходе выполнения запроса);

– просмотр с экранной страницы ресурса (статью, книгу и т.д.) в файловом виде. Форматы файлов: текстовые документы, графический образ документа (для архивных ресурсов), документ аудио- или видеодокументы (для организации электронных выставок и коллекций);

– экспертная оценка пользователем найденных абзацев;

– автоматизированная сборка в отдельный файл тематически ориентированных абзацев из различных документов электронной библиотеки вместе с их библиографическими описаниями.

V. СТРУКТУРИРОВАННОЕ ОПИСАНИЕ КОНТЕКСТНОГО ЗНАНИЯ ДЛЯ ЦЕЛЕЙ ИССЛЕДОВАНИЯ

В рамках проводимого исследования в качестве источников контекстного знания были выбраны статьи из научных журналов и публикации из средств массовой информации (СМИ). Исходя из целей исследования под микроконтекстом мы понимаем структурный текстовый элемент, содержащий как минимум одно упоминание исследуемого термина. При таком подходе под это определение попадают элементы метаданных научной публикации: название, аннотация, список ключевых слов. Кроме этого, микроконтекстом являются абзацы текста публикации, а также список пристатейной литературы. Для русскоязычных текстов также важным является наличие метаданных на английском языке, что позволяет связать русскоязычную терминологическую базу с терминологическими базами на других языках, что также является важным при сравнительном анализе

развития терминологической базы междисциплинарного научного направления «цифровая экономика» в региональном аспекте. В отличие от научных публикаций информационные сообщения из СМИ не обладают обширным набором метаданных. В большинстве из них для анализа можно использовать только название и абзацы текста.

Метаданные научной публикации вместе с текстом и списком пристатейной литературы задают вертикальную структуру контекстного знания, совокупность которых представляет собой макроконтент. Этот макроконтент является самостоятельной единицей научного информационного потока и непосредственным образом влияет на динамику развития терминологической базы данного междисциплинарного научного направления. Однако, степень влияния на развитие терминологической базы у каждого макроконтента определяется как частотой упоминания термина в макроконтенте, так и расположением его в вертикальной структуре последнего. Предлагается для каждого макроконтента дать весовые коэффициенты основным структурным его элементам (каждому элементу метаданных, абзацу текста и списку пристатейной литературы). Минимальный весовой коэффициент $K_{\text{мк}}$ имеет вхождение термина в абзац текста (микроконтент), а величина коэффициента при этом прямо пропорциональна частоте его вхождения в микроконтексты $K_{\text{мк}} \cdot n$. Далее вес коэффициентов увеличивается в порядке перечисления следующих элементов:

$K_{\text{т}}$ – вхождение термина в заглавие контекста (вес коэффициента – 5),

$K_{\text{ан}}$ – вхождение термина в аннотацию (вес коэффициента – 10),

$K_{\text{кл}}$ – вхождение термина в список ключевых слов (вес коэффициента – 100).

В предлагаемой градации важнейшим микроконтекстом, в который может входить исследуемый термин, является список ключевых слов, т.е. термин для контекста является ключевым. Вхождение термина в заглавие контекста (например, название статьи) является менее значимым, так как не всегда авторы вносят в него ключевые для статьи термины. При таком подходе можно рассчитать значение степени влияния того или иного контекста i на употребление исследуемого термина:

$$\sum K_i = \sum K_{\text{мк}} \cdot n, 5 \cdot K_{\text{т}}, 10 \cdot K_{\text{ан}}, n, 100 \cdot K_{\text{кл}}$$

Такой подход позволяет отбирать для анализа макроконтенты с высокой степенью влияния на развитие терминологической базы, производя их оценку и ранжирование.

Встречаемость исследуемого термина в списке пристатейной литературы не используется при расчёте предлагаемой величины, однако является важным показателем взаимосвязи различных контекстов по отношению к исследуемому термину и позволяет переходить к изучению других контекстов.

VI. ТЕЗАУРУС МЕЖДИСЦИПЛИНАРНОЙ НАУЧНОЙ ОБЛАСТИ

Специфика структуры тезауруса междисциплинарной научной области определяется набором научных направлений, которые участвуют в её формировании, и поэтому схематически представляет собой параллельные

вертикальные структуры, включающие выборки из тезаурусов научных дисциплин и направлений, участвующих в формировании исследуемой научной области. Один из тезаурусов является в такой ситуации основным и задает основные термины всего междисциплинарного научного направления исследований. В работе был проведён поиск научных публикаций по термин-концепту «цифровая экономика». Для этого использовалась Научная электронная библиотека (<http://elibrary.ru>). По этому же термин-концепту был произведён поиск публикаций из общественно-политического информационного потока, представленного в СМИ (данные получены из массива публикаций из более чем 1000 российских газет и журналов федерального и регионального уровней, представленных на платформе Интегрум, <https://integrum.ru>). Анализ полученных данных показал, что термин «цифровая экономика» скорее всего введён в государственном управлении (программа «Цифровая экономика Российской Федерации», утверждена распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р), причём не было дано его чёткого определения и только после этого он начал массово использоваться в научных публикациях. По нашему мнению, для «цифровой экономики» системообразующим научным направлением является не «экономика», а «цифровые технологии» («информационные технологии», «информационно-коммуникационные технологии», ИКТ). Поэтому мы считаем, что термины из этого направления будут ключевыми, а термины, используемые в других научных дисциплинах и направлениях – вспомогательными. В рамках вложенных направлений (или поднаправлений) эти термины могут быть основными. На этапе отбора цифровых информационных ресурсов нами был определён перечень наиболее часто встречающихся понятий терминологической базы междисциплинарного направления «Цифровая экономика: электронное государственное управление и умные технологии» [2], [14]. При этом были выделены основные предметные научные области: экономика, информатика, социология, образование, политология, управление. Анализ терминологической базы этих направлений показал, что термин-концепт «цифровые технологии» входит в тезаурусы четырёх из шести включённых в исследование научных областей. Он является основным в научной области «информатика». Применяя в дальнейшем онтологический подход [10], [11], [21], можно будет определить наборы основных понятий, участвующих в формировании тезауруса рассматриваемого междисциплинарного направления. Также на основе анализа результатов запросов в различных прикладных системах могут быть уточнены взаимосвязи между терминами и построены соответствующие графы.

VII. ИССЛЕДОВАНИЕ ФОРМИРОВАНИЯ ТЕРМИНОЛОГИЧЕСКОЙ БАЗЫ МЕЖДИСЦИПЛИНАРНОЙ НАУЧНОЙ ОБЛАСТИ

А. Определение основных термин-концептов

Для выявления перечня основных понятий терминологической базы междисциплинарного направления «Цифровая экономика: электронное государственное управление и умные технологии» был произведён анализ информации, полученной из открытых источников (основные массивы были получены из «Научной электронной библиотеки» и полнотекстовой базы публикаций СМИ «Интегрум»). Исходным документом была выбрана программа «Цифровая экономика Российской Федерации» (утверждена распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р), после опубликования которой наблюдается положительная динамика (взрывного характера) употребления термин-концепта «цифровая экономика» как в общественно-политическом дискурсе (СМИ), так и в научных публикациях (см. выше). Динамика на массиве полученных данных о числе публикаций, распределённых по годам приведена в таблице 1 и на рисунке 1. В программе в качестве основного направления указано создание условий для «улучшения доступности и качества государственных услуг для граждан», а среди заявленных целей выделено «создание экосистемы цифровой экономики Российской Федерации, ... в которой обеспечено эффективное взаимодействие ... государства и граждан». На основании анализа текста программы «Цифровая экономика Российской Федерации» было выделено 16 основных термин-концептов [14]. Этот список был дополнен термин-концептами «цифровые технологии» и «умные технологии», которые в ряде контекстов показали себя как синонимичные. Количественный анализ публикаций, представленных в НЭБ (eLIBRARY), показал, что данный термин используется реже, но имеет такую же динамику, как и «информационные технологии», т.е. является его подмножеством. Термин-концепт «умные технологии» связан с различными направлениями реализации программы России «Цифровая экономика», например, используется во включённой в неё по инициативе Министра программы «Умный город». Эти термины являются достаточно устоявшимися в мировом общественно-политическом и научном дискурсах (Smart Technology, Smart City).

Соотнесение указанных терминов с различными областями знания позволяет сделать вывод о междисциплинарности самого понятия «цифровая экономика» в частности и всего научного направления в целом.

Таблица. Распределение по годам научных публикаций и публикаций из СМИ по термин-концепту «цифровая экономика»

Годы	ЦЭ (eLibrary)	ЦЭ (СМИ)
2005	0	12
2006	0	7
2007	1	0

2008	5	8
2009	2	6
2010	6	8
2011	16	18
2012	15	30
2013	20	39
2014	27	48
2015	60	82
2016	148	308
2017	1268	3833

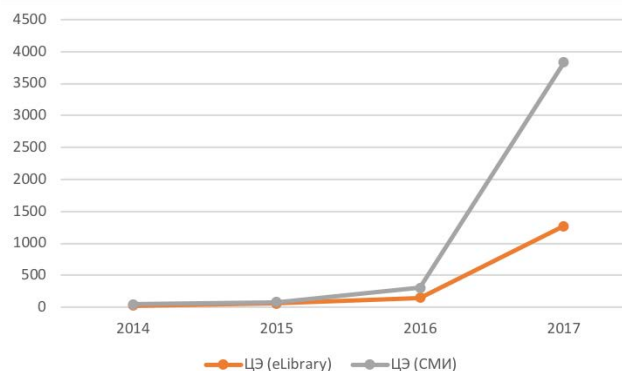


Рис. 1. Динамика распределения по годам научных публикаций и публикаций СМИ по термин-концепту «цифровая экономика»

Анализ представленного семантического ядра говорит о том, что цифровая экономика в области электронного государственного управления продолжает развитие идей и разработок, которые начали развиваться в России в последнее двадцатилетие и характеризуются устойчивыми терминами (термин-концептами): «государственное управление», «электронное управление», «электронное государство», «государственная информационная система». В контексте развития как направления «государственное управление», так и «цифровая экономика» всё чаще употребляется термин-концепт «государственные услуги» (в том числе и «муниципальные услуги»). Также для обоих направлений характерно то, что они развиваются в контексте Стратегии развития информационного общества в Российской Федерации на 2017 - 2030 годы, утвержденной Указом Президента Российской Федерации от 9 мая 2017 г. № 203 "О Стратегии развития информационного общества в Российской Федерации на 2017-2030 годы". Поэтому к этим направлениям также относятся такие термин-концепты, как «информационные технологии», «информационно-коммуникационные технологии».

В. Отбор цифровых информационных ресурсов

Для дальнейшего исследования развития терминологической базы рассматриваемого междисциплинарного направления мы произвели оценку на релевантность основным термин-концептам русскоязычных полнотекстовых цифровых электронных ресурсов. На этих ресурсах представлены: научные публикации (eLibrary, Киберленинка, Соционет, East View); научная, научно-популярная и учебно-методическая литература (ЭБС Лань, East View); а также СМИ – газеты и журналы (Интегрум, East View). Анализ результатов из eLibrary и East View указывают на то, что как в научном, так и в общественно-политическом дискурсе рассматриваемые термин-концепты начинают

появляться с начала 2000-х годов, что соответствует началу развития информационного общества в России как устойчивой общественной тенденции [14].

Для первичной оценки динамики развития терминологической базы рассматриваемого междисциплинарного направления был выполнен поиск научных публикаций в Научной электронной библиотеке (eLibrary) по термин-концептам «цифровая экономика», «электронное управление» и «электронное правительство». Выборка включила в себя публикации начиная с 2005 года. Полученные результаты были распределены по годам (рис. 2).

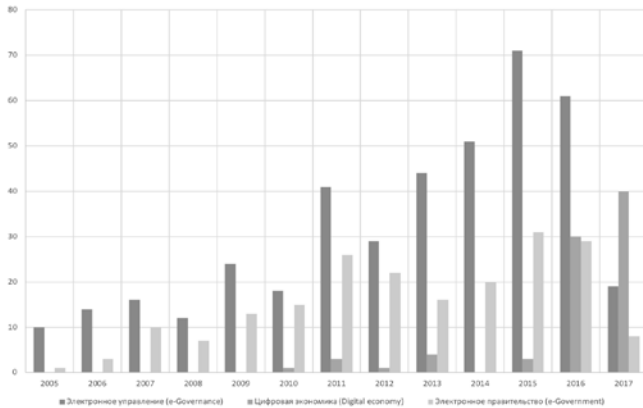


Рис. 2. Динамика публикаций по термин-концептам на электронном ресурсе "eLibrary"

Динамика публикаций показывает устойчивый рост употребления термин-концепта «электронное управление» и его более частое употребление по сравнению с термином «электронное правительство» (показатели 2017 года можно во внимание не принимать в связи с тем, что не все публикации этого года были внесены в базу данных ресурса). Начиная с 2010 года в научный дискурс начинает вводиться термин «цифровая экономика». При этом резкий всплеск его употребления приходится на 2016 и 2017 годы.

С. Применение методов контекстного поиска к задачам исследования

Задачи изучения развития терминологической базы междисциплинарной научной области могут быть решены с использованием механизмов извлечения контекстного знания и методов контекстного поиска из доступных информационных массивов. Под контекстным знанием следует подразумевать умение правильно «читать» контекст, извлекать и интерпретировать профессионально значимую информацию из любых источников. Под контекстным поиском – метод последовательного поиска фрагментов текстовых записей релевантных пользовательскому запросу в соответствии с требуемым контекстом [15]. Важным аспектом практического применения методов контекстного поиска является поддержка научно-исследовательской деятельности в новых и междисциплинарных областях. В предлагаемом исследовании подходы изучения контекстного знания являются базовыми в разрабатываемом авторами синтетическом методе, объединяющем различные подходы и инструменты Digital Humanities, для решения задач отбора цифровых информационных ресурсов, извлечения из них контекстного знания и на его анализе

уточнения терминологической базы формирующегося междисциплинарного научного направления «Цифровая экономика: электронное государственное управление и умные технологии».

В качестве инструмента поиска, экспликации и анализа данных была выбрана распределенная русскоязычная среда для научных исследований T-Libra. T-Libra – децентрализованная среда под управлением пользовательского браузера и с ориентацией на Веб-сервисы и Интернет-протоколы [21], [23], [24].

В электронную библиотеку среды T-Libra были загружены отобранные в Научной электронной библиотеке тексты по тематике «Электронное правительство и электронное управление» (за 2005, 2011, 2016 и 2017 года). Эти тексты были сгруппированы в 3 корзины (2005, 2011, 2016/17 годы).

На этом массиве были реализованы: абзацно-ориентированные запросы (ключевые слова «цифровая экономика» и «электронное правительство»), частотно-ориентированные запросы по массиву (учёт осуществлялся по словам, приведенным к форме существительных единственного числа именительного падежа). Запросы позволили выявить количество документов и релевантных тематике абзацев в этих документах.

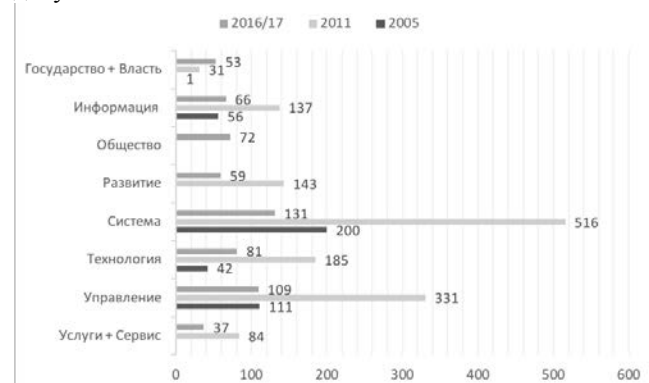


Рис. 3. Частота употребимости терминов при абсолютно-частотном запросе по массиву научных текстов «Электронное правительство и электронное управление»

Абсолютный частотно-ориентированный запрос демонстрирует динамику изменения встречаемости терминов в текстах разных лет (рис. 3). Показательно, например, что термин «система» встречается в каждой корзине в несколько раз чаще, чем термин «информация». Из этого следует, что термин «система» употребляется в нескольких разных контекстах. Один из которых – информационные системы. Видно, что интерес к темам государства, власти и услуг, сервисов появляется только в 2011 году и вырастает к 2017. Интерес к теме «Электронное правительство и электронное управление» в контексте «общество» формируется только к 2016 году.

Относительный частотный запрос производился по терминам «экономика», «власть», «технология», «государство». Его результаты позволили выявить наиболее часто встречающиеся понятия в абзацах с заданным термином и частоту встречаемости самого термина.

Анализ полученных результатов позволяет сделать вывод о возможности и корректности использования среды T-Libra для экспликации и анализа

русскоязычных текстов для исследования развития русскоязычной терминологической базы формирующегося междисциплинарного научного направления «Цифровая экономика: электронное государственное управление и умные технологии».

VIII. ВЫВОДЫ

Проведённое пилотное исследование доказало эффективность использования синтетического метода изучения контекстного знания на примере исследования развития терминологической базы междисциплинарного научного направления «Цифровая экономика: электронное государственное управление и умные технологии». При этом:

- выявлена возможность использования цифровых информационных ресурсов для пополнения распределённой информационной среды по исследуемой предметной области междисциплинарного научного направления «Цифровая экономика: электронное государственное управление и умные технологии» (основными ресурсами являются: «Научная электронная библиотека» для научных публикаций, Интегрум для СМИ и East View для научных публикаций и СМИ);

- выявлены механизмы использования распределённой среды для расширения семантического ядра (понятийной базы) предметной области междисциплинарного научного направления «Цифровая экономика: электронное государственное управление и умные технологии». Понятийная база расширена новыми термин-концептами;

- показана возможность исследования динамики развития терминологической базы предметной области междисциплинарного научного направления «Цифровая экономика: электронное государственное управление и умные технологии» с использованием цифровых информационных ресурсов на примере «Научной электронной библиотеки» (eLibrary);

- на существующем массиве документов выявлены возможности использования распределённой русскоязычной среды для научных исследований T-Libra для решения задач исследования развития терминологической базы развивающихся междисциплинарных научных областей.

Дальнейшее развитие исследования предполагает:

- пополнение текстами распределённой информационной среды;

- полнотекстовый анализ для выявления кластеров контекстного знания, соотнесенные с термин-концептами предметной области;

- формирование семантического ядра, связанного с направлением «умные технологии» в контексте развития предметной области «цифровая экономика»;

- выявление динамики развития термин-концептов предметной области междисциплинарного научного направления «Цифровая экономика: электронное государственное управление и умные технологии» с учётом их взаимосвязей.

Исследование поддержано Российским фондом фундаментальных исследований, проект № 18-011-00923.

БИБЛИОГРАФИЯ

- [1] Domingue J., Fensel D., Hendler J.A. Handbook of Semantic web Technologies. – Heidelberg; Dordrecht; London; N.Y.: Springer, 2011. 1077 p.
- [2] Kononova O., Prokudin D. Synthetic Method in Interdisciplinary Terminological Landscape Research of Digital Economy // SHS Web Conf., 50 (2018) 01082. DOI: <https://doi.org/10.1051/shsconf/20185001082>.
- [3] Saifa H., Heb Y., Fernandez M., Alani H. Contextual semantics for sentiment analysis of Twitter // Information Processing & Management. 2016. 52(1). P. 5–19. DOI: 10.1016/j.ipm.2015.01.005.
- [4] Tao F. et al. EventCube: multi-dimensional search and mining of structured and text data // Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '13). ACM, New York, NY, USA, 2013. P. 1494–1497. DOI: 10.1145/2487575.2487718.
- [5] Taylor W.P. A comparative study on ontology generation and text clustering using VSM, LSI, and document ontology models. Clemson University, 2007.
- [6] Turney P.D., Pantel P. et al. From frequency to meaning: Vector space models of semantics // Journal of Artificial Intelligence Research. 2010. 37 (1). P. 141–188.
- [7] Апанович З.Н., Марчук А.М. Новые подходы к нормализации словарей и установлению идентичности сущностей при обогащении контента научных баз знаний // XIV национальная конференция по искусственному интеллекту с международным участием КИИ-2014 (24-27 октября 2014 г., Казань, Россия): Труды конференции. В 3-т., Казань: Изд-во РИЦ «Школа». 2014. Т. 1. С. 92-100.
- [8] Бухановский А.В., Васильев В.Н. Современные программные комплексы компьютерного моделирования e-science // Известия высших учебных заведений. Приборостроение. 2010. Т. 53, № 3. С. 60-64.
- [9] Ермаков А. Е. Эксплицирование элементов смысла текста средствами синтаксического анализа-синтеза // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2003. М.: Наука. 2003.
- [10] Загорюлько Ю.А., Боровикова О.И. Построение многоязычного тезауруса предметной области средствами технологии создания порталов научных знаний // Всероссийская конференция с международным участием "Знания - Онтологии - Теории" (ЗОНТ-2011). Новосибирск, 2011.
- [11] Загорюлько Ю.А., Боровикова О.И., Кононенко И.С., Соколова Е.Г. Подход к разработке русско-английского тезауруса по компьютерной лингвистике // Труды 13 Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2011. Воронеж, 2011. С. 9-16.
- [12] Каныгин Г.В., Полтинникова М.С. Социальное знание и вопросы разработки его инструментов, Контекстно-ориентированные онтологические методы в социологии // Петербургская социология сегодня. Сборник трудов Социологического института РАН. СПб, 2015. С. 359-373.
- [13] Кононова О.В., Ляпин С.Х., Прокудин Д.Е. Синтетический метод извлечения контекстного знания в русскоязычной социально-гуманитарной сфере: комплексный подход // Информационное общество: образование, наука, культура и технологии будущего. Выпуск 1 (Труды XX Международной объединенной научной конференции «Интернет и современное общество», IMS-2017, Санкт-Петербург, 21 – 23 июня 2017 г. Сборник научных статей). СПб: Университет ИТМО, 2017. С. 52-67.
- [14] Кононова О.В., Ляпин С.Х., Прокудин Д.Е. Методы применения сетевой научной среды для изучения терминологической базы междисциплинарной научной области «цифровая экономика// Информационное общество: образование, наука, культура и технологии будущего. Выпуск 2 (Труды XXI Международной объединенной научной конференции «Интернет и современное общество», IMS-2018, Санкт-Петербург, 31 мая – 2 июня 2018 г. Сборник научных статей). СПб: Университет ИТМО, 2018. С. 37-51.
- [15] Кононова, О.В., Крутько Е.А., Ляпин С.Х. Технологии извлечения знаний на службе научно-исследовательской деятельности в вузе // Информационное общество. 2016. № 6. С. 25-37.

- [16] Куршев Е.П., Осипов Г.С., Рябков О.В., Самбу Е.И., Соловьева Н.В., Трофимов И.В. Интеллектуальная метапоисковая система // Труды международного семинара Диалог'2002. Компьютерная лингвистика и интеллектуальные технологии. М.: Наука. 2002. С. 320–330.
- [17] Ляпин С.Х., Куковякин А.В. Контекстное знание и его изучение с помощью инструментов полнотекстовой библиотеки // Научный сервис в сети Интернет: труды XVIII Всероссийской научной конференции (19-24 сентября 2016 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша. 2016. С. 240-248.
- [18] Ляпин С.Х., Куковякин А.В. Тематические коллекции полнотекстовых запросов для изучения контекстного знания (проект Humanitiana) // Сборник научных трудов XVIII Объединенной научной конференции «Интернет и современное общество». СПб.: Университет ИТМО, 2015. С. 216-224.
- [19] Осипов Г. С., Куршев Е. П., Кормалев Д. А., Трофимов И. В., Рябков О. В., Тихомиров И. А. Семантический поиск в среде интернет. Переславль-Залесский, ИПС РАН. 2003.
- [20] Осипов Г. С., Тихомиров И. А., Смирнов И. В.. Интеллектуальный поиск в глобальных и локальных вычислительных сетях и базах данных // Программные системы: теория и приложения. Труды международной конференции. Программные системы: теория и приложения. ИПС РАН, г. Переславль-Залесский, май 2004. М.: Физматлит, 2004. Т. 1. С. 21-23.
- [21] Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В. Онтологии и тезаурусы: учебное пособие. Казань, М., 2006. 157 с.
- [22] Черный А.В., Тузовский А.Ф. Развитие информационной системы организации с использованием семантических технологий // Знания–Онтологии–Теория: Матер. Всеросс. конф. с междунар. участием. – Новосибирск, 20–22 октября 2009. Новосибирск: ЗАО «РИЦ Прайс-Курьер», 2009. Т. 2. С. 52–59.

Studying the Interdisciplinary Terminological Landscape of Digital Economy with the Use of Contextual Analysis Tools

O.V. Kononova, S.Kh. Lyapin, D.E. Prokudin

Abstract — Interdisciplinary research directions are one of the main drivers of both social development and one of the main directions of transformation of scientific research focused on this development. The dynamics of the development of interdisciplinary research areas (for example, the digital economy: e-government and smart technologies; socio-ethical aspects of the digital economy; gamification in the social, scientific and educational spheres; Informatization of scientific activities) in the digital age is associated with the development of information society technologies. Also, technologies are one of the main tools that allow to quickly and efficiently conduct research on the dynamics of this development. The research is carried out on the arrays of information obtained from various digital sources, reflecting both the latest scientific publications and containing the content of social and political discourse (for example, publications in the media).

This study shows the possibility of using the developed synthetic method to solve the problems of selection of digital information resources, extraction of contextual knowledge from them and its analysis to clarify the terminological landscape of the emerging interdisciplinary scientific direction "Digital economy: e-government and smart technologies". The distributed Russian-language environment for scientific research T-Libra is used as a tool for data search, explication and analysis.

Keywords — Contextual Knowledge, Contextual Search, Knowledge Extraction, Digital Economy, e-Governance, Knowledge Research and Retrieval, Smart Technology, Distributed Network Environment, Information Resources, Interdisciplinary Scientific Directions, Terminological Landscape

REFERENCES

- [1] Domingue J., Fensel D., Hendler J.A. Handbook of Semantic web Technologies. – Heidelberg; Dordrecht; London; N.Y.: Springer, 2011. 1077 p.
- [2] Kononova O., Prokudin D. Synthetic Method in Interdisciplinary Terminological Landscape Research of Digital Economy // SHS Web Conf., 50 (2018) 01082. DOI: <https://doi.org/10.1051/shsconf/20185001082>.
- [3] Saifa H., Heb Y., Fernandez M., Alani H. Contextual semantics for sentiment analysis of Twitter // Information Processing & Management. 2016. 52(1). P. 5–19. DOI: 10.1016/j.ipm.2015.01.005.
- [4] Tao F. et al. EventCube: multi-dimensional search and mining of structured and text data // Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '13). ACM, New York, NY, USA, 2013. P. 1494–1497. DOI: 10.1145/2487575.2487718.
- [5] Taylor W.P. A comparative study on ontology generation and text clustering using VSM, LSI, and document ontology models. Clemson University. 2007.
- [6] Turney P.D., Pantel P. et al. From frequency to meaning: Vector space models of semantics // Journal of Artificial Intelligence Research. 2010. 37 (1). P. 141–188.
- [7] Apanovich Z.N., Marchuk A.M. Novye podhody k normalizacii slovarej i ustanovleniju identichnosti sushhnostej pri obogashhenii kontenta nauchnyh baz znanij // XIV nacional'naja konferencija po iskusstvennomu intellektu s mezhdunarodnym uchastiem KII-2014 (24-27 oktjabrja 2014 g., Kazan', Rossija): Trudy konferencii. V 3-t., Kazan': Izd-vo RIC «Shkola». 2014. T. 1. S. 92-100.
- [8] Buhanovskij A.V., Vasil'ev V.N. Sovremennye programnye komplekxy komp'juternogo modelirovanija e-science // Izvestija vysshih uchebnyh zavedenij. Priborostroenie. 2010. T. 53, # 3. S. 60-64.
- [9] Ermakov A. E. Jeksplicirovanie jelementov smysla teksta sredstvami sintaksicheskogo analiza-sinteza // Komp'juternaja lingvistika i intellektual'nye tehnologii: trudy Mezhdunarodnoj konferencii Dialog'2003. M.: Nauka. 2003.
- [10] Zagorul'ko Ju.A., Borovikova O.I. Postroenie mnogojazychnogo tezaurusa predmetnoj oblasti sredstvami tehnologii sozdaniya portalov nauchnyh znanij // Vserossijskaja konferencija s mezhdunarodnym uchastiem "Znanija - Ontologii - Teorii" (ZONT-2011). Novosibirsk, 2011.
- [11] Zagorul'ko Ju.A., Borovikova O.I., Kononenko I.S., Sokolova E.G. Podhod k razrabotke russko-anglijskogo tezaurusa po komp'juternoj lingvistike // Trudy 13 Vserossijskoj nauchnoj konferencii «Jelektronnye biblioteki: perspektivnye metody i tehnologii, jelektronnye kollekcii» - RCDL'2011. Voronezh, 2011. S. 9-16.
- [12] Kanygin G.V., Poltinnikova M.S. Social'noe znanie i voprosy razrabotki ego instrumentov, Kontekstno-orientirovannye ontologicheskie metody v sociologii // Peterburgskaja sociologija segodnja. Sbornik trudov Sociologicheskogo instituta RAN. SPb, 2015. S. 359-373.
- [13] Kononova O.V., Ljapin S.H., Prokudin D.E. Sinteticheskij metod izvlechenija kontekstnogo znanija v russkojazychnoj social'no-gumanitarnoj sfere: kompleksnyj podhod // Informacionnoe obshhestvo: obrazovanie, nauka, kul'tura i tehnologii budushhego. Vypusk 1 (Trudy XX Mezhdunarodnoj ob"edinennoj nauchnoj konferencii «Internet i sovremennoe obshhestvo», IMS-2017, Sankt-Peterburg, 21 – 23 ijunja 2017 g. Sbornik nauchnyh statej). SPb: Universitet ITMO, 2017. S. 52-67.
- [14] Kononova O.V., Ljapin S.H., Prokudin D.E. Metody primenija setevoj nauchnoj sredy dlja izuchenija terminologicheskoy bazy mezhdisciplinarnoj nauchnoj oblasti «cifrovaja jekonomika// Informacionnoe obshhestvo: obrazovanie, nauka, kul'tura i tehnologii budushhego. Vypusk 2 (Trudy XXI Mezhdunarodnoj ob"edinennoj nauchnoj konferencii «Internet i sovremennoe obshhestvo», IMS-2018, Sankt-Peterburg, 31 maja – 2 ijunja 2018 g. Sbornik nauchnyh statej). SPb: Universitet ITMO, 2018. S. 37-51.
- [15] Kononova, O.V., Krut'ko E.A., Ljapin S.H. Tehnologii izvlechenija znanij na sluzhbe nauchno-issledovatel'skoj dejatel'nosti v vuze // Informacionnoe obshhestvo. 2016. # 6. S. 25-37.
- [16] Kurshev E.P., Osipov G.S., Rjabkov O.V., Sambu E.I., Solov'eva N.V., Trofimov I.V. Intellektual'naja metapoiskovaja

- sistema // Trudy mezhdunarodnogo seminar Dialog'2002. Komp'juternaja lingvistika i intellektual'nye tehnologii. M.: Nauka. 2002. S. 320–330.
- [17] Ljapin S.H., Kukovjakin A.V. Kontekstnoe znanie i ego izuchenie s pomoshh'ju instrumentov polnotekstovoj biblioteki // Nauchnyj servis v seti Internet: trudy XVIII Vserossijskoj nauchnoj konferencii (19-24 sentjabrja 2016 g., g. Novorossijsk). M.: IPM im. M.V. Keldysha. 2016. S. 240-248.
- [18] Ljapin S.H., Kukovjakin A.V. Tematicheskie kollekcii polnotekstovyh zaprosov dlja izuchenija kontekstnogo znanija (proekt Humanitariana) // Sbornik nauchnyh trudov XVIII Ob'edinennoj nauchnoj konferencii «Internet i sovremennoe obshhestvo». SPb.: Universitet ITMO, 2015. S. 216-224.
- [19] Osipov G. S., Kurshev E. P., Kormalev D. A., Trofimov I. V., Rjabkov O. V., Tihomirov I. A. Semanticheskij poisk v srede internet. Pereslavl'-Zalesskij, IPS RAN. 2003.
- [20] Osipov G. S., Tihomirov I. A., Smirnov I. V.. Intellektual'nyj poisk v global'nyh i lokal'nyh vychislitel'nyh setjah i bazah dannyh // Programmnye sistemy: teorija i prilozhenija. Trudy mezhdunarodnoj konferencii. Programmnye sistemy: teorija i prilozhenija. IPS RAN, g. Pereslavl'-Zalesskij, maj 2004. M.: Fizmatlit, 2004. T. 1. S. 21-23.
- [21] Solov'ev V.D., Dobrov B.V., Ivanov V.V., Lukashevich N.V. Ontologii i tezaury: uchebnoe posobie. Kazan', M., 2006. 157 s.
- [22] Chernij A.V., Tuzovskij A.F. Razvitie informacionnoj sistemy organizacii s ispol'zovaniem semanticheskikh tehnologij // Znanija–Ontologii–Teorija: Mater. Vseross. konf. s mezhdunar. uchastiem. – Novosibirsk, 20–22 oktjabrja 2009. Novosibirsk: ZAO «RIC Prajs-Kur'er», 2009. T. 2. C. 52–59.