Исследование возможности разработки машиночитаемого каталога компьютерных программ и сред извлечения и анализа контекстного знания

О.В. Кононова[1], Д.Е. Прокудин[1,2]

[1] Университет ИТМО, [2] Санкт-Петербургский государственный университет

kononolg@yandex.ru, hogben.young@gmail.com

Аннотация

Для исследователей в современных условиях развития и тотального применения информационно-коммуникационных технологий встаёт актуальная проблема выбора эффективных средств для использования в целях исследования. Проблема обусловлена не столько огромных количеством существующего программного обеспечения, сколько отсутствием классификаций программного обеспечения (ПО) и информационных систем (ИС), обусловленными классами научно-исследовательских задач. В рамках реализуемого авторами проекта разработки подхода к исследованиям развития тематик и понятийно-терминологического аппарата междисциплинарных научных направлений рассматривались и применялись методы поиска, извлечения, уточнения, экспликации, анализа и представления контекстного знания с применением соответствующих ПО и ИС. Специфика проводимого исследования ограничивает используемые ПО и ИС задачами обработки контекстного научного знания. Были проанализированы основные типы ПО и ИС, применяемых для этих целей, выявлены основные их функциональные характеристики. В соответствии с разработанной в ходе исследования типологией контекстов и выявленными группами характеристик предлагается подход к разработке каталога ПО и ИС анализа контекстного знания с функциями выделения, классификации и экспликации научного контента. Для представления информации о ПО и ИС в каталоге предложена модель метаданных Dublin Core, что позволяет не только структурировано описать основные характеристики ПО и ИС, но и представить каталог в машиночитаемой форме, что позволяет решать задачи пополнения каталога, эффективного поиска необходимого ПО и ИС в соответствии с исследовательскими задачами, а также интеграции его в научное информационное пространство на принципах открытой науки. Также предлагается паллиативное решение для отработки корректности представления метаданных по спецификации Dublin Core и обмена метаданными по протоколу OAI-PMH.

Ключевые слова: программное обеспечение, информационные системы, классификация, каталог, контекстное знание, Dublin Core, OAI-PMH

Библиографическая ссылка: Кононова О.В., Прокудин Д.Е. Исследование возможности разработки машиночитаемого каталога компьютерных программ и сред извлечения и анализа контекстного знания // Информационное общество: образование, наука, культура и технологии будущего. Выпуск 4 (Труды XXIII Международной объединенной научной конференции «Интернет и современное общество», IMS-2020 (сборник научных статей). — СПб: Университет ИТМО, 2020. С. 42-57. DOI: 10.17586/2587-8557-2020-4-42-57

Введение

Многообразие программного обеспечения, прикладных сред и веб-ориентированных сервисов различного назначения ставит перед современными исследователями задачу выбора инструментов, которые можно эффективно использовать при проведении научных исследований. При выборе приходится ориентироваться на существующие подходы к классификации программного обеспечения (ПО) и информационных систем (ИС). Разработаны и используются как общие подходы, выделяющие самые общие классы программного обеспечения, так и частные, которые ориентированы на более детальное описание подклассов ПО, предназначенных для применения в различных прикладных областях. К наиболее общим классификациям относится, например, Классификатор программ для электронных вычислительных машин и баз данных, используемый государственными структурами в России [14]. Одним из общих подходов к классификации ПО и ИС класса Business intelligence (BI) в мировой практике является ежегодно обновляемый аналитический доклад Gartner «Infrastructure and Applications Worldwide Software Market Definitions» [31], который отражает общий подход Gartner, применяемый для оценки развития рынка программного обеспечения [26]. Подобного подхода придерживается International Data Corporation (IDC) – одна из крупнейших мировых консалтинговых компаний, ведущий поставщик информации и консультационных услуг, организатор мероприятий на рынках информационных технологий, телекоммуникаций и потребительской техники [19]. Также для классификации программного обеспечения и информационных систем может быть использована Computing Classification System (разработана Association for Computing Machinery, последняя версия была представлена в 2012 году). Авторы представляют её в качестве единого источника категорий и понятий, отражающих современное состояние областей, связанных с вычислительной техникой, информатикой и информационно-коммуникационными технологиями [21]. Широко распространены и другие подходы к классификации, которые исторически сложились из логики развития вычислительной техники и программного обеспечения [8, 9, 13].

Анализ подходов к классификации ПО и ИС позволяет выявить проблему выбора конкретного инструментария, с помощью которого исследователи как пользователи программных продуктов могут решать исследовательские и аналитические задачи. Эта проблема исходит из того, что разработчики программного обеспечения стараются встроить в свои системы максимальный набор функциональных возможностей и обеспечить тем самым выполнение целого спектра задач. Но получаемые результаты часто неравноценны. Конкретная компьютерная программа или ИС имеет свою «специализацию», то есть минимальный набор функционала, обеспечивающего выполнение более узкого спектра задач, но наиболее оптимальным и удачным образом. К тому же, как правило, существуют многочисленные аналоги, выбор между которыми будет связан с соблюдением некоторых исходных требований исследователя к системе (например, поддержка того или иного языка или реализация в ПО или ИС тех или иных групп методов). Также следует оговорить ситуации, когда информационная система не предназначалась разработчиками для обработки научной информации или анализа информации в научных целях. Но при наличии вводной, пояснительной информации или разработанного метода может быть использована в научных целях и весьма эффективно. Даже имея детальную информацию о конкретной компьютерной программе или ИС, не всегда однозначно можно классифицировать её в соответствии с существующими классификациями. Поэтому во многом как классификация, так и выявление конкретного функционала могут быть произведены или уточнены только в процессе применения ПО или ИС при решении конкретных исследовательских задач. А наличие аналогов ставит задачу создания каталогов ПО и ИС, ориентированных на решение определённого класса исследовательских задач, в соответствии разработанной классификацией.

1. Подход к поиску и отбору компьютерных программ и сред

В рамках реализуемого проекта разработки подхода (синтетический метод) к исследованиям развития тематик и понятийно-терминологического аппарата междисциплинарных научных направлений, который предполагает применение методов поиска, извлечения, уточнения, экспликации, анализа и представления контекстного знания на основе применения информационно-коммуникационных технологий, одним из основных был выбран принцип независимости синтетического метода от выбора конкретного инструментария. Специфика проводимого исследования привела к разработке типологии контекстного знания текстовой модальности, которую необходимо учитывать при выборе соответствующих ПО и ИС для обработки контекстов определённых видов (корпус, фрагмент, абзац, предложение, термин-концепт, тезаурус, мета-описание, семантическая группа, тематическая коллекция и т.д.), что позволяет использовать ПО и ИС для многоуровневого структурного анализа, повышающего эффективность решения исследовательских задач. Исходя из направленности используемых в исследовании методов на поиск, извлечение, экспликацию, анализ и представление контекстного знания, в рассмотренных общих классификациях отсутствует чёткое разделение на классы по этим укрупнённым функциям. Также по этим классификациям невозможно определить типы обрабатываемых контекстов.

Так, например, Gartner выделяет следующие сегменты рынка BI:

- средства построения хранилищ и витрин данных (Data Warehouse);
- инструменты оперативной аналитической обработки (On-Line Analytical Processing, OLAP) и прочие средства многомерного анализа;
- информационно-аналитические системы (Enterprise Information Systems, EIS) и системы поддержки и принятия решений (Decision Support Systems, DSS);
- средства интеллектуального анализа данных (Data Mining);
- инструменты конечного пользователя для выполнения запросов и построения отчетов (query and reporting tools).
- В классификации Computing Classification System рассматриваемым нами системам соответствуют следующие подклассы:
- Specialized information retrieval – Structure and multilingual text search;
- Document management and text processing – Document capture – Document searching; Document analysis.
- В отечественном Классификаторе рассматриваемым программам можно сопоставить следующие подклассы прикладного ПО:
- Поисковые системы – Программные системы поиска текстовой, графической и другой информации в локальных, корпоративных и иных хранилищах. В том числе консультационно-информационные системы поиска и просмотра информации в специализированных многоотраслевых базах данных;
- Лингвистическое программное обеспечение – Парсеры и Семантические анализаторы/Системы анализа текстов на естественных языках с выделением синтаксических структур в предложениях или выделением семантических отношений между элементами текста и общего смысла текстов;
- Системы сбора, хранения, обработки, анализа, моделирования и визуализации массивов данных – Системы бизнес-анализа (BI)/Программы, ориентированные на обработку больших объемов неструктурированных данных с целью облегчения их интерпретации, в том числе инструменты извлечения и трансформации данных (ETL), предметно-ориентированные информационные базы данных (EDW), средства аналитической обработки в реальном времени (OLAP), интеллектуального анализа данных (Data Mining), формирования отчетов, графиков, диаграмм и иных визуальных форм, поддержки принятия решений (DSS).

Для поиска и выявления ПО и ИС анализа контекстного знания с функциями выделения, классификации и экспликации научного контента для поддержки научных исследований использовались как открытые сетевые источники, так и научные публикации, в которых при проведении исследований используются подобное ПО и ИС. Анализ выявленных систем показал, что подавляющее их большинство применяются как средства лингвистического анализа текстов в различных научных исследованиях: лингвистике [4, 17], социологии [2], кибербезопасности [12], а также в междисциплинарных областях [3, 11, 18].

Также в сети Интернет можно обнаружить достаточное число каталогов лингвистических программ [1, 5, 6, 13, 30, 31, 32], которые возможно использовать в том числе и в научно-исследовательских целях.

Анализ каталогов позволяет выделить общие направления использования ПО и ИС для очерченного круга задач:

- интеллектуальный анализ текста (Text Mining);
- анализ текстов (Text Analytics/Analysis);
- поиск и извлечение информации (Information Retrieval/Extraction);
- сравнение текстов (Text Comparison);
- тематическая кластеризация/моделирование (Topic Clustering/Modelling);
- визуализация текста (Text Visualization).

В подобных каталогах отсутствуют классификаторы, в которых учитывалось бы основное их функциональное назначение и типы обрабатываемых контекстов. Поэтому они не выполняют задач выбора эффективных инструментальных средств для проведения исследования. От использования других широко представленных в сети классификаций информационных систем, в том числе по признакам «сферы применения» и «функциональность систем» было решено отказаться по тем же причинам и потому, что представленные классификации предназначены для бизнес-сообщества, оперируют бизнес-понятиями, ориентируются на круг бизнес-задач предприятия, организации или анализ рынка.

В связи с этим целью данного исследования является разработка структурированного описания ПО и ИС с использованием в качестве основных характеристик: класса ПО и ИС, основных функций и видов обрабатываемых контекстов. Также на основе структурированного описания будет исследована возможность разработки каталога ПО и ИС, который позволит исследователям решать задачу оперативного и эффективного выбора ПО и ИС, удовлетворяющих целям и задачам проводимого ими исследования. Создание каталога основывалось на необходимости решения прагматической задачи осуществить исследователю осознанный выбор необходимого и достаточного набора ПО и ИС.

2. Разработка каталога компьютерных программ и сред с функциями и сервисами извлечения и анализа контекстного знания для научных исследований

2.1. Подход к разработке классификации ПО и ИС

При формировании структуры описания ПО и ИС для представления в разрабатываемом каталоге за основу была выбрана общая направленность классификации по признаку применимости этих программ и систем для анализа контекстного знания с функциями выделения, классификации и экспликации научного контента для поддержки научных исследований.

В связи с тем, что целевой группой пользователей каталога являются ученые и преподаватели, специализирующиеся на междисциплинарных научных исследованиях и работающие с различными источниками информации и большими данными,

то в соответствии с этим были определены следующие общие и свойственные в значительной степени междисциплинарным исследованиям классы задач:
– нейросети (Neural Network);
– машинное обучение (Machine Learning);
– обработка текстов на естественном языке (Natural Language Processing);
– извлечение информации (Information Extraction);
– построение онтологий (Ontology);
– построение и анализ трендов (системы прогнозирования);
– создание и использование тезаурусов;
– тематическая классификация текстов;
– полнотекстовые базы информации;
– реферативные базы информации (только метаданные).

Этим классам задач в каталоге соответствует характеристика «тип ПО и ИС». Для последних двух типов характерны только полнотекстовые и реферативные базы информации, которые обладают собственными механизмами поиска, отбора и анализа информации.

Укрупнённые типы ПО и ИС не всегда отражают многообразие их функциональных возможностей, поэтому для более полного представления об их возможностях и рационального выбора для конкретных исследовательских целей в отдельную характеристику выделены основные функции (возможности) ПО и ИС, предназначенные для выполнения научных задач. По этому признаку выделяются следующие основные функции:
– классификация;
– прогнозирование;
– контекстный анализ;
– отбор данных по различным критериям (интеллектуальный поиск);
– автоматизированный обмен данными (метаданными);
– визуализация.

Это набор основных функций. Но при решении о включении в каталог конкретного программного продукта его анализ может выявить и другие специфические функции. Поэтому классификатор функций является расширяемым.

Предлагаемая классификация не учитывает ряд важных классов задач, не включенных в рассмотрение как выходящих за границы сугубо научно-исследовательских задач, таких как «научная коммуникация» и «вопросы управления наукой и координации научных исследований».

2.2. Виды обрабатываемых контекстов как одна из основных характеристик классифиикации ПО и ИС

Выбор конкретного ПО или ИС для использования в исследовании связан с возможностью обрабатывать те или иные виды контекстов. В рамках проводимого исследования понятие контекста понимается как независимая понятийная единица категориального аппарата, используемая в качестве основы классификации научных текстов, а также для визуализации иерархических и ассоциативных отношений между терминами. Экспликация и анализ контекстного знания при реализации проекта позволили разработать типологию контекстного знания [10], которая в дальнейшем может быть специфицирована для изучения более определённых предметных областей. Также соотнесение ПО и ИС с видами обрабатываемых контекстов позволит исследователям более рационально подходить к их выбору. Поэтому при классификации ПО и ИС предлагается включить в качестве существенной характеристики вид обрабатываемых контекстов, который соотносится с типами контекстов как некоторыми классами высшего порядка.

На основе конкретизации видов хранимых, извлекаемых и обрабатываемых (анализируемых) контекстов и в соответствии с этим ПО и ИС, отбираемые для включения в каталог, можно разделить на следующие укрупнённые категории:

– Поисковые ИС **с** возможностью обработки большого объема неформализованных текстов, возможностью обработки мультимедийной информации, которые при этом имеют ограничения в диалоге пользователя с системой, а также ограничены графематическим анализом и низкой достоверностью выявления связей (ИПС Яндекс, Google и т.п.).
– ИС, представляющие текстовые базы данных – представляют библиотечные онлайновые архивы научных публикаций и реферативные базы широкого спектра предметных областей знаний, существенно отличаются по функционалу контент-анализа (eLibrary, T-Libra, Science Direct, Scopus, WoS и др.).
– Информационно-аналитические системы, в разной степени обладающие полнотой выявления фактов, самообучением системы, возможностью обработки большого объема неформализованных текстов, возможностью обработки мультимедийной информации, уровнями семантической иерархии, уровнями автоматического логического анализа фактографической информации (Mallet, AskNet, Voyant-Tools, Tropes, Sketch Engine, CLAVIRE, RCO (Russian Context Optimizer) и др.).
– Многофункциональные ИС смешанного типа, которые обладают достоинствами и недостатками ИС, описанных выше: возможностью обработки большого объема неформализованных текстов, возможностью обработки мультимедийной информации, достоверностью выявления связей, широким спектром форматов обрабатываемых документов, но могут иметь ограничения в автоматическом уровневом семантическом анализе, не входящем в задачи компании-разработчика ИС (ABBYY Intelligent Tagger SDK, ABBYY Smart Classifier SDK; Title: PROMT Analyser и др.).

2.3. Представление каталога в машиночитаемой форме

Анализ представления информации о ПО и ИС, предназначенных для анализа контекстного знания с функциями выделения, классификации и экспликации научного контента, позволяет сделать вывод о том, что в основном подобные каталоги представляют собой статические списки или таблицы, в которых ПО и ИС либо сгруппированы по определённому признаку (например, свободно распространяемые или коммерческие; принадлежность к укрупнённым категориям функционального назначения), либо представлены в неструктурированном виде с кратким описанием возможностей и ссылками на соответствующие сайты в сети Интернет. Такое представление информации делает весьма затруднительным оперативный поиск и эффективный отбор необходимых для проведения исследования ПО и ИС с учётом основных возможностей, функционала, типов и форматов обрабатываемых контекстов.

Для целей повторного извлечения информации о программных системах и компонентах предлагается использовать их описания, как и описания документов, через представление метаданными. Так, например, González и van der Meer рассматривая стандартные форматы представления метаданных (Dublin Core, EAD, ISAD(G) и MARC), предлагают использовать расширенный набор метаданных XDC-SC (Extended Dublin Core for Software Components) схемы Dublin Core, который позволит извлекать информацию о программном обеспечении стандартными средствами поисковых систем или инструментами работы с XML [25]. Они также предполагают, такой подход может подтолкнуть к созданию сред для представления в них информации программном обеспечении. Другие исследователи предлагают не останавливаться на каком-либо одном стандарте, а разработать генеральный семантический каталог метаданных SMMC (Semantic Master Metadata Catalogue) для обеспечения взаимодействия между

существующими моделями метаданных (такими как Dublin Core, UNIMARC, MARC21, RDF/RDA и BIBFRAME) на основе реализации модели сопоставления онтологий [20]. Авторы предлагают подход к разработке семантически обогащённый экосистемы метаданных программного обеспечения SMESE (Semantic Enriched Metadata Software Ecosystem), предназначенной для поддержки конкретных распределенных приложений управление контентом. Однако, реализация такого решения является сложной задачей, решение которой возможно только на уровне консорциума.

Основываясь на общепринятых подходах для описания ПО и ИС в каталоге предлагается использовать спецификацию представления метаданных Dublin Core, в соответствии с которой основные характеристики представленных в каталоге ПО и ИС описываются соответствующими элементами основного набора метаданных Dublin Core Metadata Element Set, DCMES) [22]. При таком подходе комбинация значений этих элементов достаточно полно описывает представленные в каталоге ПО и ИС, что соответствует общему подходу использования данной спецификации к описанию различных сущностей [16, 27, 29]. Также предлагаемый подход позволяет представлять каталог в машиночитаемой форме для размещения в сетевых информационных системах со свободным доступом как исследователям, так и для осуществления автоматизированного поиска и идентификации поисковыми системами. При этом пользователи смогут осуществлять поиск по различным элементам метаданных: классам, функциям ПО и ИС или по типам обрабатываемых контекстов.

При разработке структуры записи каталога пришлось столкнуться с тем, что до сих пор спецификация Dublin Core не применялась для описания ПО и ИС. В основном в рамках данного подхода описываются различные текстовые объекты: статьи, книги, библиотечные каталожные карточки, архивные материалы, семантические модели и т.д. [29]. Поэтому в связи со спецификой описания ПО и ИС помимо основных элементов метаданных для уточнения значений характеристик были использованы квалификаторы, которые составляют второй уровень метаданных и уточняют значения элементов [7, 23].

Также для представления каталога в машиночитаемой форме был произведён выбор наиболее подходящей для этого программной платформы. При выборе мы исходили из следующих основных принципов:

– доступности – программное обеспечение с открытым кодом или некоммерческое ПО;
– популярности – наиболее известное и широко распространённое решение;
– гибкости – возможности подстраивать описание метаданных под задачу создания каталога.

Среди рассмотренных систем самой популярной в настоящее время является программная платформа DSpace (https://duraspace.org/dspace/). По данным самого авторитетного агрегатора ROAR [28] из 4725 зарегистрированных в нём репозиториев открытого доступа 1965 используют DSpace. На втором месте – EPrints (679). DSpace удовлетворяет всем выше перечисленным критериям, поэтому свой выбор мы остановили на этой платформе. Изучив документацию по представлению в DSpace метаданных в соответствии со спецификацией Dublin Core [24] и рассмотрев особенности представления использования квалификаторов, для каждого элемента каталога предлагается следующий набор метаданных:

`dc.title` — название ПО или ИС;
`dc.creator` — разработчик;
`dc.subject.classification` — основные функции (могут дополняться после анализа соотвествующих ПО и ИС);
`dc.subject.other` — вид обрабатываемого контекста;
`dc.description.abstract` — описание ПО или ИС;
`dc.publisher` — издатель (правообладатель);
`dc.contributor` — внёсший вклад (люди или организации, которые также принимали участие в

разработке ПО или ИС);
dc.date.issued — год последнего релиза;
dc.type — категории (классы ПО или ИС в соответствии с разработанной класификацией);
dc.format.mimetype — форматы обрабатываемых документов;
dc.identifier.uri — идентификатор (ссылка в сети Интернет на сайт разработчика);
dc.source.uri — источник (ссылка на веб-приложение);
dc.language — языки обрабатываемых документов;
dc.relation.isreferencedby — отношения (список публикаций по использованию ПО или ИС);
dc.coverage — поддерживаемые операционные системы;
dc.rights.license — тип лицензии.

В соответствии с предлагаемым подходом было описано более 50 ПО и ИС. Наглядное представление записи каталога в соответствии со спецификацией Dublin Core представлено на рисунке 1.

dc.title	Voyant-Tools
dc.creator	Stéfan Sinclair, McGill University; Geoffrey Rockwell, University of Alberta
dc.subject.classification	обработка отдельных документов; обработка коллекцией документов (корпус текстов); классификация; анализ интернет-страниц; частотный анализ; контекстный анализ; контекстуализация тенденций (построение трендов); визуализация данных анализа
dc.subject.other	термин; абзац; документ; коллекция документов
dc.description	Веб-ориентированная система для загрузки и анализа цифровых текстов, изучения частот и распределений терминов в документах и в коллекции документов (корпус). Представляет собой набор различных функциональных модулей. Существует локальное решение в виде приложения на JETTY
dc.publisher	Voyant-Tools
dc.contributor	Andrew MacDonald; Cyril Briquet; Lisa Goddard; Mark Turcato
dc.date.issued	2018
dc.type	обработка текстов на естественном языке
dc.format.mimetype	txt; rtf; doc; docx; pdf; zip; html; xml
dc.identifier.uri	https://voyant-tools.org
dc.source.uri	http://voyeurtools.org/voyant-server/
dc.language	Мультиязычность
dc.coverage	Web-ориентированное приложение (Web-интерфейс), Mac, Windows, JETTY server, Voyant server
dc.rights	Свободно распространяемое ПО
dc.relation. isreferencedby	Laurie J. Sampsel (2018) Voyant Tools, Music Reference Services Quarterly, 21:3, 153-157, DOI: 10.1080/10588167.2018.1496754; Welsh, Megan E. “Review of Voyant-Tools”. Collaborative Librarianship, vol. 6, no. 2, 2014, p. 96+. Academic OneFile; Sinclair, Stéfan; Rockwell, Geoffrey (2016). "Voyant Facts". Hermeneutica: Computer-Assisted Interpretation in the Humanities. Stéfan Sinclair & Geoffrey Rockwell. Retrieved 2016-12-20; Using Voyant for Text Analysis: http://voyeurtools.org/using-voyant-for-text-analysis/; Rambsy, Kenton (2016). Text-Mining Short Fiction by Zora Neale Hurston and Richard Wright. Using Voyant Tools // CLA Journal. № 59 (3): 251–258; Priestley Alexis. Voyant Tools: A Tutorial for Text Analysis: https://medium.com/@priestleyal/voyant-tools-a-tutorial-for-text-analysis-df265d85d214;
dc.coverage	Любая
dc.rights.license	Creative Commons Attribution 4.0 International (CC BY 4.0)

Рис. 1. Описание ИС Voyant-Tools в соответствии со спецификацией Dublin Core

2.4. Реализация и использование каталога

Несмотря на выбор программной платформы DSpace для машинной реализации каталога установка и настройка этой системы не является тривиальной задачей, что не позволило сразу же реализовать это решение. В связи с этим для первоначального тестирования в качестве паллиативного решения было выбрано свободно распространяемое программное обеспечение с открытым кодом Open Journal Systems (OJS, https://pkp.sfu.ca/ojs/), представляющее собой платформу полного издательского цикла для издания электронных журналов. Это решение уже было применено для машиночитаемого представления тезауруса в рамках реализуемого проекта. OJS более проста в установке и настройке и может работать на большинстве виртуальных хостингов. Эта система обладает всем необходимым функционалом: поддерживает формат представления метаданных Dublin Core, позволяет осуществлять поиск по метаданным, предоставляет открытый доступ к информации, выполняет роль провайдера по протоколу OAI-PMH. В экспериментальных целях в установленную систему OJS (http://ojs.iculture.spb.ru/index.php/thesauri) были введены описания нескольких единиц ПО и ИС из отобранных. Для контроля корректности отображения метаданных и проверки работы протокола OAI-PMH была использована инсталляция системы Open Harvester Sustems (OHS, https://pkp.sfu.ca/ohs/), являющаяся агрегатором метаданных по протоколу OAI-PMH. На рисунке 2 приведён пример представления метаданных в OHS.

FIELD	VALUE
Title	Voyant-Tools
Creator	Sinclair, Stéfan Rockwell, Geoffrey
Subject	обработка отдельных документов; обработка коллекцией документов (корпус текстов); классификация; анализ интернет-страниц; частотный анализ; контекстный анализ; контекстуализация тенденций (построение трендов); визуализация данных анализа — термин; абзац; документ; коллекция документов
Description	Веб-ориентированная система для загрузки и анализа цифровых текстов, изучения частот и распределений терминов в документах и в коллекции документов (корпус). Представляет собой набор различных функциональных модулей. Существует локальное решение в виде приложения на JETTY.
Publisher	Информационные системы
Contributor	Andrew MacDonald Cyril Briquet Lisa Goddard Mark Turcato
Type	info:eu-repo/semantics/article info:eu-repo/semantics/publishedVersion — обработка текстов на естественном языке
Identifier	http://ojs.iculture.spb.ru/index.php/systems/article/view/5
Source	Информационные системы; Компьютерные программы и среды с функциями и сервисами извлечения и анализа контекстного знания для научных исследований
Language	rus
Relation	http://ojs.iculture.spb.ru/index.php/systems/article/view/5/2 http://ojs.iculture.spb.ru/index.php/systems/article/downloadSuppFile/5/2
Coverage	Web-ориентированное приложение (Web-интерфейс); Mac; Windows; JETTY server; Voyant server — —
Rights	(c) 2018 Voyant-Tools https://creativecommons.org/licenses/by/4.0/

Рис. 2. Описание ПО и ИС схемой Dublin Core, полученное в агрегаторе OHS по протоколу OAI-PMH

Предлагаемый подход к представлению каталога ПО и ИС в машиночитаемой форме также позволяет экспортировать метаданные в другие форматы представления для размещения в различных информационных системах и агрегаторах метаданных.

Использование протокола обмена метаданными OAI-PMH реализует возможность интеграции каталога в информационное пространство научных исследований. Исследователи могут не только осуществлять поиск в каталоге, но и создавать собственные информационные системы и агрегировать в них информацию из каталога. Также использование такой платформы как DSpace позволит использовать её в качестве агрегатора и собирать информацию о использовании представленных в каталоге ПО и ИС из различных ресурсов, поддерживающих представление метаданных в спецификации Dublin Core по протоколу OAI-PMH. Например, это могут быть научные публикации, в которых рассматривается применение того или иного ПО и ИС в конкретных научных целях (рис. 3).

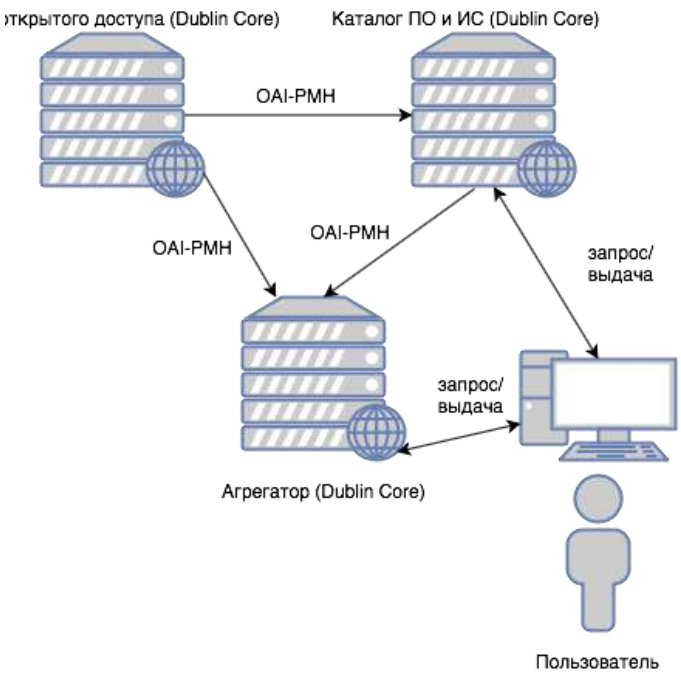

Рис. 3. Использование машиночитаемого каталога в распределённой информационной среде

Такая модель распределённой информационной среды позволит представить в одном информационном пространстве не только описания ПО и ИС, но и информацию об их применении, что предоставит исследователям методическую поддержку для более рационального выбора и эффективного использования того или иного инструментария в своих научных целях.

Выводы

Проведённое исследование показало, что в мировой практике отсутствует единый общепринятый подход к классификации ПО и ИС, предназначенных для анализа контекстного знания с функциями выделения, классификации и экспликации научного контента, в котором бы учитывались виды обрабатываемых контекстов.

Также было выявлено, что отсутствуют разработки по представлению каталогов ПО и ИС в машиночитаемом виде на основе описания их в формате метаданных.

Разработанный подход к представлению каталога ПО и ИС, предназначенных для анализа контекстного знания с функциями выделения, классификации и экспликации научного контента, на базе Dublin Core обеспечивает:

- интеграцию в каталог разработанной типологии контекстов, представляющей собой существенную характеристику, которая является основанием выбора ПО и ИС для проведения конкретных исследований;
- создание машиночитаемого каталога с использованием стандартного свободно распространяемого программного обеспечения (например, OJS, DSpace);
- эффективный поиск и отбор необходимых ПО и ИС для целей исследования в соответствии с основными характеристиками, описанными в тегах Dublin Core, используя стандартные поисковые механизмы;
- открытый доступ к элементам каталога как для пользователей, так и для автоматизированного индексирования;
- автоматизированный обмен по протоколу OAI-PMH для агрегации мета описаний каталога в других информационных системах.

Предлагаемое паллиативное решение (OJS) в будущем предполагается заменить на информационную систему, построенную на свободно распространяемом программном обеспечении DSpace. Параллельно продолжится работа по наполнению каталога описаниями ПО и ИС, которые могут быть использованы для анализа контекстного знания с функциями выделения, классификации и экспликации научного контента. Помимо этого, будет производиться отбор научных публикаций, описывающих результаты исследований, полученные с использованием размещённых в каталоге ПО и ИС.

Большой потенциал дальнейшего использования разрабатываемого каталога в рамках учебной деятельности – магистры образовательной программы «Цифровые технологии умного города» направления подготовки «Прикладная информатика» будут использовать его для выбора технических средств, необходимых для реализации своих исследовательских проектов. Каталог является одной из составных частей разрабатываемого учебно-методического комплекса «Технологии извлечения и интеллектуального анализа данных в научных исследованиях», направленного на формирование исследовательских и аналитических компетенций магистрантов. На основе разрабатываемого УМК будет модернизирован учебный курс «Информационные технологии в научной деятельности».

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

Литература

[1] 4 Free and Open Source Text Analysis Software. https://www.softwareadvice.com/resources/easiest-to-use-free-and-open-source-text-analysis-software/ (дата обращения: 17.02.2020).

[2] Видясова Л.А., Тенсина Я.Д. Результаты семантического анализа текстов в СМИ о развитии «умных городов» в России // Государство и граждане в электронной среде. Выпуск 2 (Труды XXI Международной объединенной конференции «Интернет и современное общество, IMS-2018, Санкт-Петербург, 30 мая - 2 июня 2018 г. Сборник научных статей). — СПб: Университет ИТМО, 2018. С. 112-117. DOI: 10.17586/2541-979X-2018-2-112-117.

[3] Гегер А.Э., Чупахина Ю.А., Гегер С.А. Компьютерные программы для анализа качественных и смешанных данных // Петербургская социология сегодня. 2015. № 6. С. 374-388. http://www.pitersociology.ru/ru/node/407 (дата обращения: 17.02.2020).
[4] Иванова А.А. Риторика военных компьютерных игр (по результатам контент-анализа) // Компьютерная лингвистика и вычислительные онтологии. 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. 166 – 178. DOI: 10.17586/2541-9781-2019-3-166-178.
[5] Каталог лингвистических программ и ресурсов в Сети / Составитель С.В. Логичев. 2006. https://rvb.ru/soft/catalogue/index.html (дата обращения: 17.02.2020).
[6] Кузнецов К.И. Обзор систем извлечения данных из неструктурированных текстов. 2013. http://www.pullenti.ru/(X(1)S(ngdeikpifqat0ccmnoqanfz3))/CompetitorPage.aspx?AspxAutoDetectCookieSupport=1 (дата обращения: 17.02.2020).
[7] Квалификаторы Dublin Core (Дублинского ядра) // RUSMARC, российская версия UNIMARC. Российская национальная библиотека. http://www.rusmarc.info/soft/dcq.html (дата обращения: 17.02.2020).
[8] Классификация программного обеспечения / Алексеев Е.Г., Богатырев С.Д. Информатика. Мультимедийный электронный учебник. http://inf.e-alekseev.ru/text/Klassif_po.html (дата обращения: 17.02.2020).
[9] Классификация программного обеспечения ЭВМ / Самсонова О.В. Информатика: учебное пособие. http://tpt.tom.ru/umk/informat/uchebnik/klass.htm (дата обращения: 17.02.2020).
[10] 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/j1/article/view/882/844 (дата обращения: 17.02.2020).
[11] Кравченко Ю.А. Задачи семантического поиска, классификации, структуризации и интеграции информации в контексте проблем управления знаниями // Известия ЮФУ. Технические науки. 2016. №7 (180). С. 5-18. DOI: 10.18522/2311-3103-2016-7-518.
[12] Лаврентьев А.М., Смирнов И.В., Соловьев Ф.Н., Суворова М.И., Фокина А.И., Чеповский А.М. Анализ корпусов текстов террористической и антиправовой направленности // Вопросы кибербезопасности. 2019. № 4(32). С. 54-60. DOI: 10.21681/2311-3456-2019-4-54-60.
[13] Основы информатики и вычислительной техники: учебно-практическое пособие / Морозевич А.Н. и др. Под общ. ред. А.Н. Морозевича. М-во образования Респ. Беларусь, Белорус. гос. экон. ун-т. Минск, БГЭУ. 2005. 221 с.
[14] Приказ Минкомсвязи РФ от 31.12.2015 N 621 «Об утверждении классификатора программ для электронных вычислительных машин и баз данных» (в ред. Приказов Минкомсвязи РФ от 01.04.2016 N 134, от 30.07.2019 N 422). URL: https://normativ.kontur.ru/document?moduleId=1&documentId=345157#h74 (дата обращения: 17.02.2020).
[15] Программы лингвистического анализа и обработки текста. http://asknet.ru/analytics/programms.htm (дата обращения: 17.02.2020).
[16] Федотов А.М., Леонова Ю.В. Требования к прототипу системы управления информационными ресурсами в распределенных информационных системах поддержки научных исследований // Вычислительные технологии. 2018. Т. 23, № 5. С. 82-109. DOI: 10.25743/ICT.2018.23.5.008.
[17] Чжан П., Захаров В. П. Компьютерная визуализация русской языковой картины мира // Компьютерная лингвистика и вычислительные онтологии. 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. 92 –105. DOI: 10.17586/2541-9781-2019-3-92-105.

[18] Чугунов А.В., Кабанов Ю.А. «Электронное государство» как междисциплинарная научная область: наукометрический анализ // Государство и граждане в электронной среде. 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. 11 – 24. DOI: 10.17586/2541-979X-2019-3-11-24.

[19] Andsbjerg R., Vesset D. IDC's Worldwide Software Taxonomy, 2018: Update. https://www.idc.com/getdoc.jsp?containerId=US44835319 (дата обращения: 17.02.2020).

[20] Brisebois R., Abran A., Nadembega, A. A Semantic Metadata Enrichment Software Ecosystem (SMESE) Based on a Multi-Platform Metadata Model for Digital Libraries // Journal of Software Engineering and Applications. 2017. № 10. P. 370-405. DOI: 10.4236/jsea.2017.104022.

[21] Computing Classification System. URL: https://dl.acm.org/ccs (дата обращения: 17.02.2020).

[22] DCMI Metadata Terms // Dublin Core Metadata Initiative. https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ (дата обращения: 17.02.2020).

[23] DCMI Qualifiers // Dublin Core Metadata Initiative. https://www.dublincore.org/specifications/dublin-core/dcmes-qualifiers/ (дата обращения: 17.02.2020).

[24] DSpace/dublin-core-types.xml at master DSpace // DSpace. GitHub. https://github.com/DSpace/DSpace/blob/master/dspace/config/registries/dublin-core-types.xml (дата обращения: 17.02.2020).

[25] González R., Van Der Meer K. Standard Metadata Applied to Software Retrieval // Journal of Information Science. 2004. Vol. 30(4). P. 300–309. DOI: 10.1177/0165551504045850.

[26] Infrastructure and Applications Worldwide Software Market Definitions. Gartner Dataquest Guide, 2002. http://smartshore.us/Infrastructure_Market_trends_2003.pdf (дата обращения: 17.02.2020).

[27] Noor S., Shah L., Adil M., Gohar N., Saman G.E., Jamil S., Qayum F. Modeling and representation of built cultural heritage data using semantic web technologies and building information model // Computational and Mathematical Organization Theory. 2018. P. 1-24. DOI: 10.1007/s10588-018-09285-y.

[28] Registry of Open Access Repositories. http://roar.eprints.org (дата обращения: 17.02.2020).

[29] SUNScholar/Metadata/By Function // Libopedia. https://wiki.lib.sun.ac.za/index.php/SUNScholar/ Metadata/By_Function (дата обращения: 17.02.2020).

[30] Text Analysis, Text Mining, and Information Retrieval Software. https://www.kdnuggets.com/software/text.html (дата обращения: 17.02.2020).

[31] Text Mining Software. https://www.capterra.com/text-mining-software/ (дата обращения: 17.02.2020).

[32] Text mining, text analytics & content analysis with free open source software. https://www.opensemanticsearch.org/doc/analytics/textmining (дата обращения: 17.02.2020).

[33] Woodward A., Anderson R., Biscotti F., Contu R., Gupta N., Hunter E., Hare J., Bhullar B., Dayley A., Roth C., Swinehart H., Dsilva V., Wurster L., Poulter J., Palanca T., Deshpande S., Pang C., Abbabatulla B., Warrilow M., Dharmasthira Y., Kostoulas J. Market Definitions and Methodology: Software. 2019. https://www.gartner.com/en/documents/3906823/market-definitions-and-methodology-software (дата обращения: 17.02.2020).

Study of Developing a Machine-readable Catalog of Computer Programs and Tools for Extracting and Analyzing Contextual Knowledge

O. Kononova [1], D. Prokudin [1,2]

[1] ITMO University, [2] Saint-Petersburg State University

For researchers in the modern conditions of development and total application of information and communication technologies, there is an urgent problem of choosing effective means to use for research purposes. The problem is caused not so much by the huge amount of existing software, but by the lack of classifications of software and information systems, due to the classes of research tasks. In the framework implemented by the authors of the project-development approach to research the development of the thematic and terminological apparatus interdisciplinary scientific fields considered and applied the methods of search, extraction, clarification, explication, analysis, and presentation of contextual knowledge, applying software and information systems. The specifics of the research limit software and information systems used to the tasks of processing contextual scientific knowledge. The main types of software and information systems used for these purposes were analyzed, and their main functional characteristics were identified. In accordance with the typology of contexts developed in the course of the study and the identified groups of characteristics, an approach is proposed to develop a catalog of software and information systems analysis of contextual knowledge with the functions of allocation, classification, and explication of scientific content. To provide information about software and information systems in the catalog of the proposed metadata model is Dublin Core, which allows not only structured to describe the main characteristics of software and information systems, but also to present the catalog in a machine-readable form that allows solving problems of replenishment of the catalog effective search software and information systems the necessary and is in accordance with the research tasks, and integrate it into the scientific information space on the principles of open science. We also offer a palliative solution for testing the correctness of the presentation of metadata according to the Dublin Core specification and the exchange of metadata via the OAI-PMH Protocol.

Keywords: software, information systems, classification, catalog, contextual knowledge, Dublin Core, OAI-PMH

Reference for citation: Kononova O., Prokudin D. Study of developing a machine-readable catalog of computer programs and tools for extracting and analyzing contextual knowledge // Information Society: Education, Science, Culture and Technologies of the Future. Vol. 4 (Proceedings of the XXII International JointScientific Conference «Internet and Modern Society», IMS-2020, St. Petersburg, June 17-20, 2020). - St. Petersburg: ITMO University, 2020. P. 42 – 57. DOI: 10.17586/2587-8557-2020-4-42-57

Reference

[1] 4 Free and Open Source Text Analysis Software. https://www.softwareadvice.com/resources/easiest-to-use-free-and-open-source-text-analysis-software/ (access date: 17.02.2020).

[2] Vidiasova L., Tensina I. Results of the Semantic Analysis of Texts in Mass Media on the Development of «Smart Cities» in Russia // The State and Citizens in the Electronic Environment. Vol. 2 (Proceedings of the XXI International Joint Scientific Conference. Internet and Modern Society., IMS-2018, St. Petersburg, May 20 - June 2, 2018). – St. Petersburg: ITMO University, 2018. P. 112-117. DOI: 10.17586/2541-979X-2018-2-112-117 (In Russian).

[3] Geger A.E., Tchupakhina Y.A., Geger S.A. Computers programs for the qualitative and mixed data analysis. St. Petersburg Sociology Today. 2015. № 6. P. 374-388. http://www.pitersociology.ru/ru/node/408 (access date: 17.02.2020). (In Russian).
[4] Ivanova A.A. Rhetoric of wargames (results of the content analysis) // Computer Linguistics and Computing Ontologies. Vol. 3 (Proceedings of the XXII International Joint Scientific Conference «Internet and Modern Society», IMS-2019, St. Petersburg, June 19-22, 2019). - St. Petersburg: ITMO University, 2019. P. 266 – 278. DOI: 10.17586/2541-9781-2019-3-266-278 (In Russian).
[5] Katalog lingvisticheskikh programm i resursov v Seti / Sostavitel' S.V. Logichev. 2006. https://rvb.ru/soft/catalogue/index.html (access date: 17.02.2020). (In Russian).
[6] Kuznetsov K.I. Obzor sistem izvlecheniya dannykh iz nestrukturirovannykh tekstov. 2013. http://www.pullenti.ru/(X(1)S(ngdeikpifqat0ccmnoqanfz3))/CompetitorPage.aspx?AspxAutoDetectCookieSupport=1 (access date: 17.02.2020). (In Russian).
[7] Kvalifikatory Dublin Core (Dublinskogo yadra) // RUSMARC, rossiyskaya versiya UNIMARC. Rossiyskaya natsional'naya biblioteka. http://www.rusmarc.info/soft/dcq.html (access date: 17.02.2020). (In Russian).
[8] Klassifikatsiya programmnogo obespecheniya / Alekseev E.G., Bogatyrev S.D. Informatika. Mul'timediynyy elektronnyy uchebnik. http://inf.e-alekseev.ru/text/Klassif_po.html (дата обращения: 17.02.2020). (In Russian).
[9] Klassifikatsiya programmnogo obespecheniya EVM / Samsonova O.V. Informatika: uchebnoe posobie. http://tpt.tom.ru/umk/informat/uchebnik/klass.htm (access date: 17.02.2020). (In Russian).
[10] 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 (access date: 17.02.2020). (In Russian).
[11] Kravchenko Yu.A. Information's semantic search, classification, structuring and integration objectives in the knowledge management context problems // izvestiya SFedU. engineering sciences. 2016. №7 (180). P. 5-18. DOI: 10.18522/2311-3103-2016-7-518 (In Russian).
[12] Lavrent'ev A.M., Smirnov I.V., Solov'ev F.N., Suvorova M.I., Fokina A.I., Chepovskiy A.M. Analysis of corpus of extremist texts and unlawful texts // Voprosy kiberbezopasnosti. 2019. № 4(32). P. 54-60. DOI: 10.21681/2311-3456-2019-4-54-60 (In Russian).
[13] Osnovy informatiki i vychislitel'noy tekhniki: uchebno-prakticheskoe posobie / Morozevich A.N. i dr. Pod obshch. red. A.N. Morozevicha. M-vo obrazovaniya Resp. Belarus', Belorus. gos. ekon. un-t. Minsk, BGEU. 2005. 221 p. (In Russian).
[14] Prikaz Minkomsvyazi RF ot 31.12.2015 N 621 «Ob utverzhdenii klassifikatora programm dlya elektronnykh vychislitel'nykh mashin i baz dannykh» (v red. Prikazov Minkomsvyazi RF ot 01.04.2016 N 134, ot 30.07.2019 N 422). https://normativ.kontur.ru/document?moduleId=1&documentId=345157#h74 (access date: 17.02.2020). (In Russian).
[15] Programmy lingvisticheskogo analiza i obrabotki teksta. http://asknet.ru/analytics/programms.htm (access date: 17.02.2020). (In Russian).
[16] Fedotov A.M., Leonova Y.V. Requirements for the prototype of the information resources management system in distributed information systems for the support of scientific research // Computational technologies. 2018. V. 23. № 5. P. 82-109. DOI: 10.25743/ICT.2018.23.5.008 (In Russian).
[17] Zhang P., Zakharov V. P. Computerized visualization of the Russian language picture of the world // Computer Linguistics and Computing Ontologies. Vol. 3 (Proceedings of the XXII International Joint Scientific Conference «Internet and Modern Society», IMS-2019, St. Petersburg, June 19-22, 2019). - St. Petersburg: ITMO University, 2019. P. 92 – 105. DOI: 10.17586/2541-9781-2019-3-92-105 (In Russian).

[18] Chugunov A.V., Kabanov Y. Electronic Governance" As an Interdisciplinary Scientific Field: Scientometrics Analysis // The State and Citizens in the Electronic Environment. Vol. 3 (Proceedings of the XXII International Joint Scientific Conference «Internet and Modern Society», IMS-2019, St. Petersburg, June 19-22, 2019). – St. Petersburg: ITMO University, 2019. P. 11 – 24. DOI: 10.17586/2541-979X-2019-3-11-24 (In Russian).
[19] Andsbjerg R., Vesset D. IDC's Worldwide Software Taxonomy, 2018: Update. https://www.idc.com/getdoc.jsp?containerId=US44835319 (access date: 17.02.2020).
[20] Brisebois R., Abran A., Nadembega, A. A Semantic Metadata Enrichment Software Ecosystem (SMESE) Based on a Multi-Platform Metadata Model for Digital Libraries // Journal of Software Engineering and Applications. 2017. No. 10. P. 370-405. DOI: 10.4236/jsea.2017.104022 (In Russian).
[21] Computing Classification System. https://dl.acm.org/ccs (access date: 17.02.2020).
[22] DCMI Metadata Terms // Dublin Core Metadata Initiative. https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ (access date: 17.02.2020).
[23] DCMI Qualifiers // Dublin Core Metadata Initiative. https://www.dublincore.org/specifications/dublin-core/dcmes-qualifiers/ (access date: 17.02.2020).
[24] DSpace/dublin-core-types.xml at master DSpace // DSpace. GitHub. https://github.com/DSpace/DSpace/blob/master/dspace/config/registries/dublin-core-types.xml (access date: 17.02.2020).
[25] González R., Van Der Meer K. Standard Metadata Applied to Software Retrieval // Journal of Information Science. 2004. Vol. 30(4). P. 300–309. DOI: 10.1177/0165551504045850.
[26] Infrastructure and Applications Worldwide Software Market Definitions. Gartner Dataquest Guide, 2002. http://smartshore.us/Infrastructure_Market_trends_2003.pdf (access date: 17.02.2020).
[27] Noor S., Shah L., Adil M., Gohar N., Saman G.E., Jamil S., Qayum F. Modeling and representation of built cultural heritage data using semantic web technologies and building information model // Computational and Mathematical Organization Theory. 2018. P. 1-24. DOI: 10.1007/s10588-018-09285-y.
[28] Registry of Open Access Repositories. http://roar.eprints.org (access date: 17.02.2020).
[29] SUNScholar/Metadata/By Function // Libopedia.https://wiki.lib.sun.ac.za/index.php/SUNScholar/ Metadata/By_Function (access date: 17.02.2020).
[30] Text Analysis, Text Mining, and Information Retrieval Software. https://www.kdnuggets.com/software/text.html (access date: 17.02.2020).
[31] Text Mining Software. https://www.capterra.com/text-mining-software/ (access date: 17.02.2020).
[32] Text mining, text analytics & content analysis with free open source software. https://www.opensemanticsearch.org/doc/analytics/textmining (access date: 17.02.2020).
[33] Woodward A., Anderson R., Biscotti F., Contu R., Gupta N., Hunter E., Hare J., Bhullar B., Dayley A., Roth C., Swinehart H., Dsilva V., Wurster L., Poulter J., Palanca T., Deshpande S., Pang C., Abbabatulla B., Warrilow M., Dharmasthira Y., Kostoulas J. Market Definitions and Methodology: Software. 2019. https://www.gartner.com/en/documents/3906823/market-definitions-and-methodology-software (access date: 17.02.2020).