

Обогащённая сеть проблем как ядро метаданных электронной библиотеки

А. П. Гагарин*, И. А. Филимонов

ФГБОУ ВО «Московский авиационный институт (национальный исследовательский университет)», г. Москва, Российская Федерация

Адрес: 125993, Российская Федерация, г. Москва, Волоколамское шоссе, д. 4

* gagarin_ay@outlook.com

Аннотация

Статья направлена на совершенствование объектного поиска в сети проблем электронной библиотеки. Поиск производится «вручную» на экране компьютера, на котором показывается специальная разновидность графа знаний, представляющего как документы, хранящиеся в библиотеке, так и относящиеся к ним метаданные. Одна часть узлов графа изображает хранящиеся документы в то время, как другая часть представляет технические и научные проблемы, обсуждаемые в документах, фазы решения этих проблем, места публикаций, если таковые имели место, и авторов. Рёбра графа отражают различные отношения между этими сущностями. Различаются три уровня метаданных. Верхний уровень состоит из кластеров, которые образованы узлами и рёбрами и соответствуют конкретным по своему содержанию темам, распознаваемым в библиотечном контенте. Следующие два уровня имеют словесную природу. Метки узлов, не обязательные сами по себе, образуют второй уровень метаданных. Метка представляет собой краткое номинативное словосочетание, которое присваивается узлу, когда он включается в библиотеку. Метка содействует объектному поиску, который начинается с распознавания кластеров по когнитивному отображению их конфигураций в памяти пользователя. Поиск продолжается путём перемещения в пространстве графа и изменения масштаба, в каком он показывается на экране компьютера. При достижении определённого масштаба метки становятся видимыми и начинают влиять на поиск. Третий уровень метаданных лежит вне пространства графа и представлен таблицей. Путём выделения узла, который заинтересовал пользователя, в таблице показывается строка, в которой пользователь может получить такие сведения об узле, как библиографическую запись документа или характеристику проблемы. Этот процесс даёт информацию, достаточную, чтобы принять решение о завершении или продолжении поиска. В обоих случаях пользователь получает полный доступ к требуемым документам, хранящимся в библиотеке с помощью специального браузера. Нахождение конкретного документа оказывается значимым, но не единственным исходом пользования библиотекой. Во многих случаях к библиотеке обращаются, чтобы получить сведения о состоянии некоторой научной или технической области в сжатом и, по возможности, обогащённом виде. Предлагается обогащение метаданных путём расширения номенклатуры узлов графа, делающее метаданные библиотеки более информативными и, следовательно, более пригодными для сценариев эксплуатации библиотеки, указанных выше. Рассмотрены трудности, возникающие при обогащении метаданных, и сформулированы соответствующие рекомендации. Доклад содержит пример библиотечной коллекции, сформированной согласно этим рекомендациям.

Ключевые слова: электронная библиотека, архив документов, контент библиотеки, метаданные библиотеки, объектный поиск, граф знаний, проблема, сеть проблем, инновационный цикл

Авторы заявляют об отсутствии конфликта интересов.

Для цитирования: Гагарин А. П., Филимонов И. А. Обогащённая сеть проблем как ядро метаданных электронной библиотеки // Современные информационные технологии и ИТ-образование. 2021. Т. 17, № 4. С. 860-870. doi: <https://doi.org/10.25559/SITITO.17.202104.860-870>

© Гагарин А. П., Филимонов И. А., 2021



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Enriched Problem Network as a Core of the Metadata in a Digital Library

A. P. Gagarin*, I. A. Filimonov

Moscow Aviation Institute (National Research University), Moscow, Russian Federation

Address: 4 Volokolamskoe shosse, Moscow 125993, Russian Federation

* gagarin_ay@outlook.com

Abstract

This issue contributes to advancing of the entity search in problem networks of a digital library. The search is performed manually in a window, that visualizes some kind of knowledge graph, representing the library content involving the relevant metadata. A subset of the graph nodes depicts documents the library contains. The other part of nodes represents technological and scientific problems, being discussed in the documents, phases of problem-solving, sites of publication if any, and authoring. Edges of the graph depict diverse relations of the connected entities. Three levels of the metadata are discernable. The upper level is constituted of clusters built from nodes and edges and corresponding to particular semantic topics recognizable in the library content. The next two levels are of verbal nature. Optional labels of the nodes build the second level of library metadata. The label is a succinct nominative word composition assigned to a node when it was being included into the library. The label assists the entity search that starts from recognizing of clusters upon a cognitive map of their shapes in the user's memory. The search is continued by moving in the graph space and adjusting the scale. At some scale, the node labels grow visible and amend and rule the search. The third metadata level lies outward of the graph space and is represented by a table. Through selection of a node being of interest, the corresponding table line is displayed, and the user can read full information about the node, such as bibliographic record of the document or problem characterization. This process provides the decisive information for finishing or continuing of the search. In both cases, the user receives full access to the pertinent documents stored in the library by a special browser. Finding of a specific document often appears to be a valuable but not the single outcome of using a library. In multiple cases one turns to a library for observing a collection of sources in some scientific and technological area and get the knowledge in an accumulated and possibly generalized form. An enrichment of the problem network in the core of library metadata is proposed, making the library metadata more informative and hence more suitable for the mode of library application mentioned above. Challenges emerging during the metadata enrichment are considered and corresponding recommendations are formulated. The issue contains a sample of a library collection built pursuant to the given recommendations.

Keywords: digital library, document archives, library content, library metadata, entity search, knowledge graph, problem, problem network, innovation cycle

The authors declare no conflict of interest.

For citation: Gagarin A.P., Filimonov I.A. Enriched Problem Network as a Core of the Metadata in a Digital Library. *Sovremennye informacionnye tehnologii i IT-obrazovanie = Modern Information Technologies and IT-Education*. 2021;17(4):860-870. doi: <https://doi.org/10.25559/SITITO.17.202104.860-870>



Введение

Первичным и массово реализуемым сценарием пользования любым хранилищем информации является получение доступа к его фондам, в случае библиотеки – к хранимым документам. Доступ может быть классифицирован по переборной поисковой активности, требуемой от пользователя-«читателя».

В частности, доступ может быть прямым, который воспринимается пользователем-«читателем» как неделимая операция, требующая от него задания точного (для данной библиотеки) поискового признака и возвращающая искомый документ либо уведомление об отсутствии искомого. В этом случае в пределах отдельной операции доступа поисковая активность от читателя не предусматривается. Общеизвестным примером могут служить библиотеки модулей в компьютерных системах.

На другом полюсе подобной классификации находятся системы с объектным поиском [1]. В этом случае роль библиотеки состоит в предоставлении читателю возможности полного или частичного перебора элементов библиотечного фонда и «опознания» среди них искомых объектов на основе поисковых признаков, находящихся в когнитивной памяти [2], не обязательно предъявленных библиотеке заранее, перед началом поиска, и не обязательно оформляемых словесно. Примером может служить специализированная электронная библиотека – индивидуализированный архив научно-технической документации Поиск-УМ, в котором для объектного поиска читателю предоставляется визуальная модель сети научно-технических проблем [3], дополненная сетью инновационных циклов технических продуктов, связанных с показываемыми проблемами, и сетью документов архива.

Эта сеть визуализируется редактором графов Gephi¹ [4, 5], считывающим с файла представление сети в формате Graph Exchange XML Format (GEXF) [6, 7]. Редактор графов строит элементы сетей в своей памяти в форме объектов и показывает их в своих окнах в двух формах: в виде размеченного раскрашенного графа и в виде таблицы, отображающей атрибуты объектов.

В данной статье предлагается дополнить поисковую сеть системы Поиск-УМ и представить её на языке UML в качестве спецификации возможного расширения метамодели любой электронной библиотеки, отвечающей требованиям стандарта² ГОСТ Р 7.0.100-2018 на состав библиографической записи. Эта спецификация может быть рекомендована как основа развития функций электронной библиотеки в направлениях объектного поиска, реферирования и обобщения знаний. Научно-технические библиотеки и персонализированные

библиотеки (архивы) смогут предоставить дополнительный сценарий пользования, а именно, сценарий оперативного освоения обобщённых и структурированных знаний из некоторой области науки или техники путём чтения метаданных библиотеки, в которых эти знания накапливаются.

Электронные библиотеки и их метаданные

Теория и практика электронных библиотек как разновидности информационных систем имеет продолжительную и богатую историю. Одной из первых теоретических моделей известность получила Модель S5 [8-10], постулирующая, что каждой электронной библиотеке присущи пять компонентов: потоки, структуры, пространства, сценарии и сообщества. Модель NESTOR [11] концептуально позиционирует архив как частный случай электронной библиотеки. Указывая, что для архива характерно ограниченное распространение и возможная невозпроизводимость объектов хранения.

В [12] была предложена онтология документирования, в которой анализ научного дискурса предлагалось проводить в понятиях «предпосылка», «заключение», «вклад», «дискуссия», «мотивация», «оценка» и «сценарий». Идея анализа структуры научного дискурса с целью отображения его в метаданные восходит к последним годам XX века, когда появилась теория структурной риторики (Rhetorical structure theory) [13], продолжившая своё развитие в первом десятилетии нового века [14, 15]. К тем же годам восходит идея моделирования научного и технического прогресса на основе сети проблем³, продолжаемая в данной статье.

В последнее десятилетие укрепилось понимание того, что для электронных библиотек командный словесный интерфейс отнюдь не недостаточен. В [16] метаданные хранилища научных и инженерных публикаций были построены как онтология, выраженная на языках⁴ RDFS/OWL [17, 18].

В области электронных библиотек стали применяться графы знаний (Knowledge Graphs – KG).

В [19] можно найти, что KG состоит из онтологии, описывающей некоторую концептуальную модель, и конкретных данных, в которых соблюдаются ограничения, накладываемые онтологией. Там же формулируются требования к ORKG (Open Research KG) – разновидности KG, отличающейся подбором сценариев. К этим сценариям относятся: «Произвести обзор поля исследований», «Найти связанные работы», «Оценить релевантность», «Извлечь релевантную информацию», «Выдать рекомендуемые статьи», «Добиться глубокого понимания», «Воспроизвести результаты».

¹ Gephi – The Open Graph Viz Platform. USA, 2019. [Электронный ресурс]. URL: <https://gephi.org> (дата обращения: 17.09.2021).

² ГОСТ Р 7.0.100-2018 Система стандартов по информации, библиотечному и издательскому делу. Библиографическая запись. Библиографическое описание. Общие требования и правила составления: национальный стандарт РФ: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 3 декабря 2018 г. № 1050-ст: введен впервые: дата введения 2019-07-01 / подготовлен ИТАР-ТАСС, ФГБУ «РГБ», РНБ. М.: Стандартинформ, 2020; ГОСТ Р 7.0.96-2016 Система стандартов по информации, библиотечному и издательскому делу. Электронные библиотеки. Основные виды. Структура. Технология формирования: национальный стандарт РФ: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 8 декабря 2016 г. № 2003-ст: введен впервые: дата введения 2017-07-01 / подготовлен ФГБУ «РГБ», РНБ, ПБ, ГПНТБ России, БЕН РАН, ВИНТИ РАН. М.: Стандартинформ, 2018.

³ Gagarin A.P. Simulation of Science and Technology Progress // Proceedings of the VI International Conference “System Identification and Control Problems” (SICPRO’07). Article number: 4215. М.: V.A. Trapeznikov Institute of Control Science, 2007.

⁴ RDF Current Status [Электронный ресурс] // W3C, 2021. URL: https://www.w3.org/standards/techs/rdf#w3c_all (дата обращения: 17.09.2021); OWL Web Ontology Language Current Status [Электронный ресурс] // W3C, 2021. URL: <http://www.w3.org/standards/techs/owl> (дата обращения: 17.09.2021).



В качестве примера можно указать KG академии Майкрософт [20, 21], который связывает хранящиеся статьи по цитированию, авторам, грантам, журналам и ключевым словам.

Функции электронных библиотек поддерживаются технологиями извлечения данных, баз данных и автоматического анализа текста, которые значительно развились за последние 20 лет [22]. Следует также отметить, что объектный поиск стал ведущей формой поиска в хранилищах информации [1], [23], [24].

Развитие электронных библиотек соприкоснулось с библиотечным делом, в котором давно существовала развитая система понятий и технологий, частично опирающихся на применение автоматизированной обработки информации. Были разработаны соответствующие международные стандарты, адаптированные в России⁵. В течение последнего десятилетия появились стандарты по тематике электронных библиотек, в частности, относящиеся к их метаданным, которые необходимо учитывать в новых разработках.

При всём разнообразии современных электронных библиотек научный работник и инженер продолжают нуждаться в электронной библиотеке, которая, адаптируясь в ходе эксплуатации к тематике работы владельца и одновременно выступая как архив его работ, позволяет ему оперативно, без отрыва от творческого труда получать справки о состоянии и тенденциях развития интересующей его области науки и техники, а также проводить анализ накапливаемых в библиотеке информационных материалов.

Объектная структура метаданных электронной библиотеки

В соответствии со стандартом⁶ фонд электронной библиотеки включает в себя «первичные объекты» (храняемые документы и издания) и описательные метаданные. К ним примыкают структурные и административные метаданные. На Рис. 1 представлена спецификация совокупности метаданных, позволяющих представить научную и техническую проблематику конкретной коллекции документов, структуру артефактов и систему понятий области науки и техники, состояние которой отражено в данной коллекции. Спецификация выполнена как диаграмма классов в нотации языка UML и предназначается как основа проектирования поисковых и навигационных функций программного обеспечения библиотеки⁷.

Класс Document представляет первичные объекты библиотечной коллекции. Состав атрибутов соответствует ГОСТ⁸. Объекты этого класса связываются между собой в соответствии с библиографическими записями в документах.

Класс Problem представляет научную или техническую проблему, сформулированную или выявленную в документе библиотеки. Объекты классов Approach, Proposition, Attempt, Evaluation, Resolution содержат в атрибуте Specification информацию об этапах решения проблемы, как детально рассмотрено в [3]. Ассоциации Yield указывают возможные переходы в соответствии с историей решения. От каждой научной статьи идут связи к одному или нескольким объектам Problem. У каждого такого объекта образуется конфигурация связей и объектов, «кластер» - носитель метаданных исходной статьи. Кластеры этого типа образуют сеть проблем метаданных библиотечной коллекции.

Объекты класса Artefact и связанные с ними объекты классов Research, ОКР, Consum и Assessment образуют сеть жизненных циклов артефактов – технических изделий, в которых реализуются решения научных и технических проблем данной коллекции. Связи transmit выражают передачу результатов решения проблемы объекту Research (научные исследования) или непосредственно объекту ОКР (опытно-конструкторские работы). Проблемы могут инициироваться не только от документов, но и непосредственно от объектов Research. Концепция жизненного цикла технических изделий применена к библиотечным метаданным в [16].

Классы Issue, Publication и Periodic представляют объекты публикационной деятельности, содержащие в общем случае несколько документов библиотеки. Объекты этих классов связываются с площадками, в среде которых осуществляется публикация. Они представлены классами Forum, Conference и session. Объекты этих классов создают инфраструктуру, позволяющую наглядно упорядочить документы во времени и, косвенно, в пространстве.

Согласно принципам наследования классов все классы, наследующие класс Node, имеют атрибуты Label и Specification. Label – это метка узла, представляющая объект класса при его визуализации в виде графа. Атрибут Specification предназначен для хранения текстов, детально описывающих состояние объекта. Эти тексты образуют вербальный уровень метаданных, распределённых в соответствии семантикой классов. Необходимо отметить, что для этапов решения проблемы и элементов жизненного цикла артефактов наследование классу Node не показано, чтобы не загромождать диаграмму.

В своей совокупности конфигурации объектов указанных классов образуют сетевое ядро метаданных, которое может быть показано в окне соответствующего редактора как граф знаний, связанных с данной коллекцией электронной библиотеки.

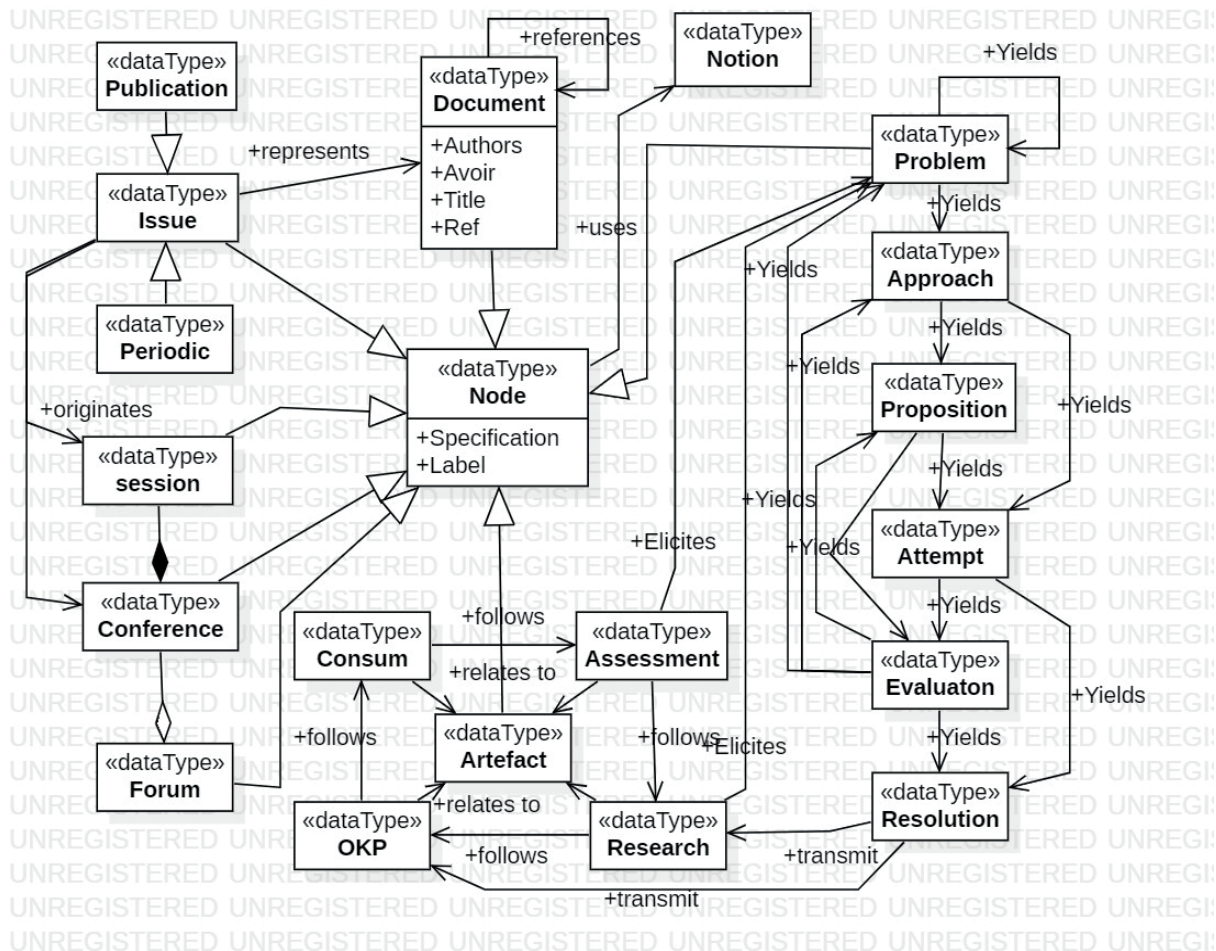
⁵ ГОСТ Р ИСО 23081-1-2008 Система стандартов по информации, библиотечному и издательскому делу. Процессы управления документами. Метаданные для документов. Часть 1. Принципы = ISO 23081-1:2006 Information and documentation – Records management processes – Metadata for records – Part 1: Principles: национальный стандарт РФ: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 13 ноября 2008 г. № 310-ст: введен впервые: дата введения 2009-07-01 / подготовлен ВНИИДАД, М.: Стандартинформ, 2020.

⁶ ГОСТ Р 7.0.96-2016 Система стандартов по информации, библиотечному и издательскому делу. Электронные библиотеки. Основные виды. Структура. Технология формирования: национальный стандарт РФ: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 8 декабря 2016 г. № 2003-ст: введен впервые: дата введения 2017-07-01 / подготовлен ФГБУ «РГБ», РНБ, ПБ, ГПНТБ России, БЕН РАН, ВИНТИ РАН. М.: Стандартинформ, 2018.

⁷ OMG Unified Modelling Language (OMG UML). Version 2.5.1. [Электронный ресурс] // OMG, 2021. URL: <http://www.omg.org/spec/UML/2.5.1/PDF> (дата обращения: 17.09.2021).

⁸ ГОСТ Р 7.0.100-2018 Система стандартов по информации, библиотечному и издательскому делу. Библиографическая запись. Библиографическое описание. Общие требования и правила составления: национальный стандарт РФ: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 3 декабря 2018 г. № 1050-ст: введен впервые: дата введения 2019-07-01 / подготовлен ИТАР-ТАСС, ФГБУ «РГБ», РНБ. М.: Стандартинформ, 2020.





Р и с. 1. Спецификация метаданных электронной библиотеки

F i g. 1. Digital library metadata specification

Объект класса Problem является источником путей, которые могут ветвиться, сливаться и завершаться объектом Resolution, из которого могут исходить связи к узлам Problem, начинающим новые пути, но, как правило, относящимся к другому кластеру. Заикливание не допускается.

На пути от объекта класса Problem к объекту класса Resolution могут сколько угодно раз повторяться объекты Problem, Approach, Proposition, Attempt, Evaluation. Объект Evaluation должен следовать за объектом Attempt. Путь от объекта Problem до объекта Resolution может быть редуцирован до одного ребра (то есть, промежуточные узлы отсутствуют). Могут сливаться объекты Approach и Proposition, Proposition и Attempt, Evaluation и Resolution. Все случаи редукции пути определённым образом отражаются в спецификациях узлов. Если спецификации не компенсируют и не объясняют редуцию пути, то модель следует считать недоработанной или некачественной.

Функция администрирования метаданных

Администрирование охватывает сценарии ввода (пополнения библиотеки), корректировки и удаления элемента хранения. Сценарий ввода распадается на ветви в зависимости от характера вводимого документа. Различаются научно-исследовательские, обзорные и рекламно-оценочные публикации. Если библиотека играет также роль архива физического лица или организации, в библиотечный фонд могут входить неопубликованные материалы владельца.

Корректировка может возникнуть при поступлении нового издания имеющегося документа. Обработка этого поступления зависит от политики библиотеки: новое издание может вводиться как самостоятельная единица хранения и связываться с предыдущим изданием соответствующим ребром (отношением). Или же новое издание не вводится в библиотеку, а



метаданные предыдущего издания изменяются, чтобы отразить контент нового издания.

Центральным действием всех указанных сценариев является создание узла в сети метаданных. Предпочтительно создавать узел через графическое представление сети, так как при этом естественным образом задаётся положение узла относительно других узлов в пространстве графа. Введённый узел выделяется («селектируется»). В открывшейся таблице пользователь может присвоить узлу значения его атрибутам.

Шаги сценария ввода документа X:

- создание узла «Document», представляющего базовые метаданные для документа X,
- формирование сетевого кластера, содержащего дополнительные метаданные для документа X,
- обеспечение консистентности преобразованной сетевой модели метаданных.

Далее эти шаги рассмотрены более подробно.

При вводе узла «Document» пользователь выполняет следующие действия:

- создаёт узел в сети,
- присваивает атрибуту Descriptor значение Document
- заносит в атрибуты узла названия документа, авторов, место и год издания и прочие (стандартизированные) библиографические данные,
- присваивает атрибуту Avoir сетевой адрес документа,
- присваивает атрибуту Specification в качестве значения полную или сокращённую аннотацию, если она предусмотрена в документе,
- обрабатывает каждую библиографическую запись, входящую в состав библиографии документа (если она имеется).

Если документ, соответствующий библиографической записи, уже входит в библиотечный фонд, то введённый документ связывается с ним специальным отношением, которое показывается ребром графа. В противном случае он вводится в библиотеку, как описано выше, но формирование сетевого кластера откладывается.

Формирование кластера для введённого узла включает в себя следующие действия:

- выявление в документе поставленных или решаемых научных и технических проблем с введением для каждой выявленной проблемы узла «Problem»; формулировка проблемы присваивается атрибуту Specification,
- выявление в документе подходов к решению выявленных проблем, предложений по их решению, попыток решения и их результатов с введением, соответственно, узлов «Approach», «Proposition», «Attempt», «Evaluation», «Resolution»; соответствующие формулировки вводятся в атрибут Specification,
- установление между введёнными узлами отношения следования, которое показывается рёбрами графа,
- ввод узла для каждого понятия, существенно используемого в текстах метаданных; для абстрактных понятий вводится узел «Notion», а для технических артефактов – узлы «Artefact», «OKP», «Consum», «NIR»; определения вводятся в атрибут Specification.

Обеспечение консистентности преобразованной сетевой модели метаданных заключается в выявлении в сетевой моде-

ли узлов, близких или тождественных по смыслу введённым. Тожественные узлы объединяются с сохранением отношений, или они связываются соответствующим отношением, которое показывается ребром специального вида (спиралью), а для близких вводятся обобщающие узлы.

Проблемы отражения контента электронной библиотеки в её метаданных

Как каталоги классических научно-технических библиотек, метаданные электронных библиотек отражают предметное содержание каждой хранимой публикации в её названии, ключевых словах и аннотации (в последней в той мере, в какой автор или авторы пожелают или сумеют это содержание отразить). Аннотация предпочтительно берётся из самой публикации. Место и время опубликования, обычно содержащееся в библиографическом паспорте и входящее в метаданные, не относятся к предмету публикации, но полезны при оценке актуальности и оригинальности опубликованных исследований, восстановлении истории и контекста их развития.

Очевидно, что углубленное отражение контента в метаданных библиотеки может быть осуществлено за счёт переработки аннотаций и списка ключевых слов, введения новых элементов в структуру метаданных и совершенствования её визуального представления. При этом обнаруживаются проблема определения целесообразного объёма метаданных (проблема граничных условий) и проблема возникновения интерпретаций хранимого документа, искажающих авторский замысел.

Проблема объёма вряд ли может быть решена в общем случае. С одной стороны, объём метаданных не должен превышать объём контента, поскольку вместе с метаданными растут затраты на их выработку и падает скорость поиска и просмотра, соответственно теряются преимущества работы с метаданными, а не прямо с контентом. С другой стороны, вряд ли целесообразно ограничивать формирование в составе метаданных полезных комментариев и обобщений, особенно при наличии средств автоматизации, ускоряющих и упрощающих эту работу.

Переработка оригинальных текстов контента для включения в метаданные может оказаться вынужденной мерой, когда структура и стиль оригинального документа не соответствуют архитектуре метаданных. Например, может оказаться затруднительным единообразно отразить в метаданных статьи об инженерных разработках, естественно-научных исследованиях, исторические и аналитические обзоры, оценочные и критические работы, заметно отличающиеся по организации и характеру изложения. Переработка документов ведёт к интерпретации содержания, искажающей или скрывающей замысел их авторов для пользователей библиотеки, граничащей с нарушением авторских прав. В то же время, девиантное толкование контента может выступать как развитие идей его авторов, естественное проявление творческого импульса специалистов, вовлечённых в процесс формирования метаданных. Дозволенная степень девиантного толкования контента в метаданных, очевидно, зависит от политики библиотеки. Она выше для индивидуализированного архива и ниже для публичных библиотек. Общим решением является, видимо, соответствующее маркирование элементов метаданных, которые



дополняют или изменяют оригинальный документ и подводят к его интерпретации, отличной от оригинальной, а также выделение таких документов в отдельный документ-рецензию, автором которого уже будет специалист, формирующий метаданные. Такая рецензия может сохраняться в составе метаданных или включаться в состав контента библиотеки (получая, в частности, свои метаданные).

Подходы к визуализации метаданных могут быть классифицированы по соотношению графики и вербализации в пространстве визуального представления (в окне экрана компьютера) и по степени когнитивности. Вербализация тем выше, чем большая площадь занята словами на известных пользователю языках, то есть имеющих для него смысл. Графика может оцениваться по суммарной площади графем (фигур), уменьшенных до различимости. Минимальная когнитивность у законченных по смыслу самостоятельных высказываний, рисунков и графических символов, имеющих устойчивое общеизвестное толкование. Среднюю когнитивность имеют слова, словосочетания и изображения, не имеющие в текущем контексте однозначного толкования. Чем больше вариантов толкования, тем выше когнитивность. Высокую степень когнитивности имеют изображения, которые не могут быть однозначно описаны, но вызывают у наблюдателя однозначную поведенческую реакцию, например, узнавания лица конкретного человека.

Типовыми вариантами визуализации являются: таблицы и графики, диаграммы различного рода, графы. Диаграммы и графы могут содержать в своём пространстве надписи, то есть, быть в большей или меньшей степени вербализованными.

Особое место занимает ветвящийся гипертекст, у которого узлы содержат текст, а рёбра выполняют единственную смысловую функцию – организацию дискурса.

При выборе графа становятся актуальными проблемы выбора плотности, соотношения размеров узлов и расстояния между узлами, формы и цвета узлов и рёбер, наличие словесных меток у узлов и рёбер. Сомнительно, что существует универсальное решение, пригодное для всех пользователей и сценариев работы с библиотекой. Обилие вариантов визуализации и их равноправие, может быть, кажущееся, вызывает рост требований к средствам визуализации. Варианты визуализации должны определяться параметрами средств визуализации, обеспечивающими выбор наилучших в ходе интенсивной практики. Один из таких вариантов рассмотрен в следующем разделе.

Электронная библиотека-архив на основе редактора графов Cytoscape

Примерная реализация предложенной спецификации проведена с применением редактора графов Cytoscape [25]. Она демонстрирует возможное решение отмеченных выше проблем. Спецификация конкретизируется, в частности, производится выбор значков и цветовой палитры для графического представления метаданных, для навигации по графу существенно используется масштабирование.

Номенклатура узлов и их графическое представление показаны на Табл.1.

Таблица 1. Номенклатура и графическое представление объектов метаданных
Table 1. Nomenclature and graphical representation of metadata objects

Наименование	Внутренний дескриптор	Назначение	Графическое представление
Артефакт	Artefact	Артефакт	Белый прямоугольник
Выпуск	Issue	Выпуск публикации	Эллипс
Документ	Avoir	Документ,	Синий треугольник
Заключение	Evaluation	Оценка проведённой реализации	Эллипс
Исследование	Research	НИР, исследование	Голубоватый прямоугольник
Локация	Conference	Конференция, семинар, симпозиум	Зеленоватый ромб
ОКР	OKP	Опытно-конструкторские работы	Спряmlённый квадрат
Оценка	Assessment	Оценка попытки или предложения	Спряmlённый квадрат
Периодическое издание	Periodic	Периодическое издание	Ромб
Площадка	Forum	Организатор встреч и публикатор («локация»)	Косой прямоугольник
Понятие	Notion	Понятие	«Галка»
Подход	Approach	Подход, выбранный для решения проблемы	Оранжевый эллипс
Попытка	Attempt	Попытка, план или модельная реализация принятого решения	Оранжевый диск
Потребление	Consum	Потребление	Голубоватый прямоугольник
Предложение	Proposition	Предложение, принципиальное решение по реализации выбранных эвристик	Вертикальный эллипс
Проблема	Problem	Проблема	Красный диск
Публикация	Publication	Публикация	Треугольник стороной вверх
Решение	Resolution	Заключение о завершении цепочки проблем и передачи результатов артефакту	Зелёный диск
Сессия	session	Сессия (внутри конференции)	Эллипс



На Табл. 2 показаны атрибуты объекта метаданных и, соответственно, узла графа, представляющего метаданные.

Т а б л и ц а 2. Атрибуты узла метаданных
Table 2. Metadata Node Attributes

Атрибут	Назначение	Примечание
Author	Автор документа, список авторов через запятую	В материале, разработанном как замечание пользователем архива, Author может опускаться
Avoir	URL	Для документов в локальном архиве компьютера http://127.0.0.1:3001/путь_до_файла/имя_файла.pdf
Descriptor	тип узла	Используется для различения цвета, формы и габаритов узла (через таблицу стилей)
label	метка	Показывается возле узла, в зависимости от типа узла может скрываться, показываться над узлом (умолчание), под узлом или на фоне узла
Ref	Библиографические реквизиты документа, кроме авторов и наименования	Для экспорта в библиографические списки, создаваемые пользователем Архива, содержит наименование издания, год издания и прочее.
Selected	да/нет	Отражает или задаёт селекцию узла
Specification	Развёрнутое описание узла	
SUID	Порядковый номер узла в сети	Создается автоматически, не удаляется

Сочетание графа и таблиц в визуальном представлении метаданных вызвано намерением минимизировать вербальную информацию в поле графа, чтобы она не отвлекала от объектного поиска по конфигурациям. Однако полностью избавиться от неё оказывается невозможным. При масштабе, достаточно крупном, чтобы без труда селективировать узлы и рёбра манипулятором (мышью), возникает необходимость точно различать элементы простых, а следовательно, повторяющихся структур, например, таких, как ветвление (при количестве ветвей, более 4). Для этого предложено использовать метки – строки с рекомендуемой длиной не более 10 литер, возможно, с пробелами. То есть, метка – это аббревиатура или сочетание от одного до трёх коротких слов, желательное, воспринимаемых как мнемоника, понятная конкретному пользователю.

Такое соглашение о метках определило размеры изображений узлов и расстояний между узлами с учётом следующих требований:

- узел не должен доминировать над меткой, при масштабе, обеспечивающем нормальное чтение метки;
- узлы должны быть различимы по форме и цвету при масштабе, при котором метки перестают быть видимыми;
- метки соседних узлов не должны сливаться.

В результате установлено, что размер узла может варьировать от 2 до 4 высоты литер в шрифте, в котором представлены метки.

Решение различать типы узлов метаданных при их визуализации позволяет повысить узнаваемость конфигураций узлов при объектном поиске. Узлы различаются как формой, так и размером. Формы выбраны простейшие: круг, овал, прямоугольник, треугольник – чтобы обеспечить узнаваемость при уменьшении. Цвет используется как дополнительный признак различения ввиду его недоступности для лиц с ограниченным восприятием цвета.

Между объектами метаданных определены отношения ассоциации, следования и эквивалентности, которые в графе изображаются прямыми линиями или дугами обычной толщины, двойной толщины и спиральной линией соответственно.

Поисковые и аналитические сценарии взаимодействия оператора с электронной библиотекой

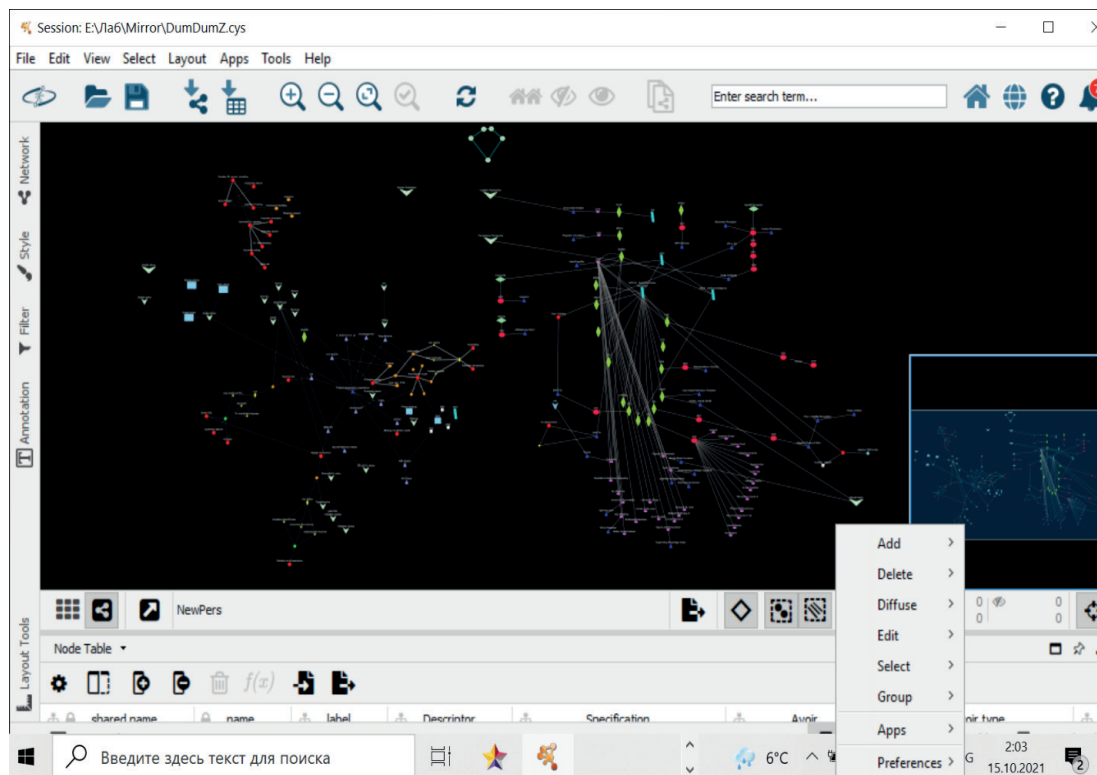
На Рис. 2 показаны метаданные коллекции документов, собранных по теме «Электронная библиотека». Поиск в общем случае начинается с кластеров узлов. Искомый кластер опознаётся визуально по конфигурации входящих в него узлов и рёбер. Внутри кластера выбор определяется меткой, формой и цветом узла. Выбранный узел селективируется, и показывающийся в Таблице узлов текст спецификации позволяет принять решение о завершении или продолжении поиска. Текст найденного документа выводится вызовом специальной функции-браузера.

В правой части Рис. 2 видны локации, в которых публикуются документы: зеленая дуга конференций, сиреневые эллипсы сессий, наклонные тонкие прямоугольники – сообщества. В левой части рисунка – кластеры нескольких документов.

Поддерживаются три сценария работы пользователя: поиск, обзор и администрирование метаданных. Редактирование может быть вызвано как новыми поступлениями в библиотеку, так и уточнением и комментированием образов хранящихся документов.

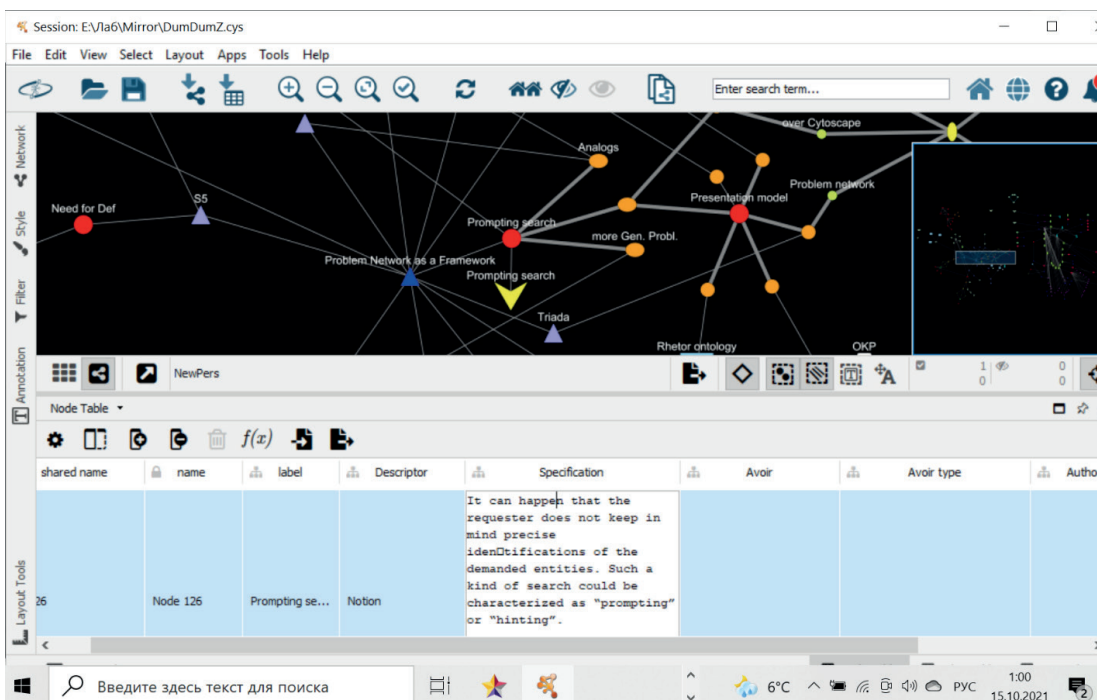
Поиск осуществляется пользователем и состоит из шагов, каждый из которых начинается навигацией по представлению метаданных в окне системы в виде графа. Шаг поиска заканчивается выбором некоторого узла и обращением к табличному представлению его атрибутов. На основании анализа этих атрибутов пользователь принимает решение закончить поиск или перейти к следующему его шагу. Из таблицы пользователь может открыть найденный документ. Навигация осуществляется сдвигами графа в окне и изменением масштаба. Пользователь как бы планирует, снижаясь, над плоскостью графа. Пока масштаб крупный, пользователь ориентируется по конфигурациям узлов и рёбер. При большем увеличении пользователь получает возможность различать и учитывать метки при узлах.





Р и с. 2. Крупномасштабный вид графа

F i g. 2. Large scale view of the graph



Р и с. 3. Отображение статьи Problem Network as a Framwork в сети проблем

F i g. 3. Displaying Problem Network as a Framwork Article in the Network of problems



Поиск имеет объектный характер, не предполагает формулирования поисковых признаков. Объектами поиска во время навигации по графу являются конфигурации узлов и рёбер с учётом или без учёта меток. При анализе табличного представления метадаанных объектом поиска выступает сам хранимый документ, точные реквизиты (наименование документа, авторы, год и место выпуска и другое) которого находятся в таблице как атрибуты текущего или смежного с ним узла.

Сценарий обзора отличается от сценария поиска тем, что навигация по графу не заканчивается выбором и открытием документа. Если обзор не претендует на глубину, достаточная информация может быть получена из табличного представления узлов графа метадаанных. Анализ табличного представления метадаанных сменяется новым шагом навигации, начиная с того же уровня масштабирования графа или более крупного. На Рис. 3 показан граф, отображающий результат поиска: метадаанные статьи и саму статью [3].

Узел-треугольник синего цвета представляет статью, красный кружок – проблему, жёлтый, коричневый и зелёный – предложение, попытку и заключение, соответственно. Рёбра графа показывают связь фаз решения между собой в порядке их следования во времени, а также связь со статьёй. Узлы снабжены метками, состоящими из одного-двух слов. Под окном, в котором представлен граф, располагается так называемая «Таблица узлов», показывающая атрибуты селектированного узла. В данном примере селектирован узел Notion Prompting Search, и Таблица узлов показывает её формулировку.

Важную роль на шагах поиска играет возможность масштабирования и сдвига графа. Сначала просматриваются кластеры на большом «удалении» (пример на Рис. 2), затем субкластеры

– детали кластеров при увеличенном масштабировании. При дальнейшем увеличении масштаба читателю становятся доступны метки и цвета узлов, как на Рис. 3.

Иной сценарий состоит в том, что оператор-читатель ограничивается работой с метадаанными, с их наиболее информативной частью – спецификациями, заключёнными в Таблицах узлов, и не прибегает к вызову полных текстов документов. Этот сценарий позволяет, в частности, оперативно аккумулировать информацию, заключённую в метадаанных для генерации обзоров. Очевидно, что данный сценарий заключает в себе значительный потенциал как интегративного, так и аналитического использования информационных ресурсов электронных библиотек. Поскольку степень формализации метадаанных библиотеки, вообще говоря, выше степени формализации публикуемых статей и докладов, эти метадаанные являются более благоприятным материалом для автоматизации аналитических и прогнозных работ в конкретных областях науки и техники.

Заключение

Кроме решения задачи поиска, метадаанные библиотеки обеспечивают поддержание творческой осведомлённости читателя в моделируемых темах. Особенно это свойство может быть полезно для работников, вынужденных или склонных одновременно заниматься несколькими темами. Непосредственная доступность текстов публикация делает избыточным формирование спецификаций путём копирования этих текстов. Однако целесообразно стремиться к тому, чтобы тексты спецификаций по пути решения образовывали связный дискурс.

References

- [1] Balog K. Entity-Oriented Search. *The Information Retrieval Series*. Vol. 39. Springer Cham; 2018. 351 p. (In Eng.) doi: <http://doi.org/10.1007/978-3-319-93935-3>, 2018
- [2] Siew C.S.Q., Wulff D.U., Beckage N.M., Kenett Y.N., Meštrović A. Cognitive Network Science: A Review of Research on Cognition through the Lens of Network Representations, Processes, and Dynamics. *Complex*. 2019; 2019:24. (In Eng.) doi: <https://doi.org/10.1155/2019/2108423>
- [3] Gagarin A.P., Filimonov I.A. Problem Network as Access Gate into a Document Repository. *Sovremennye informacionnye tehnologii i IT-obrazovanie = Modern Information Technologies and IT-Education*. 2020; 16(3):582-597. (In Russ., abstract in Eng.) doi: <https://doi.org/10.25559/SITITO.16.202003.582-597>
- [4] Heymann S., Grand B.L. Visual Analysis of Complex Networks for Business Intelligence with Gephi. *2013 17th International Conference on Information Visualisation*. IEEE Press, London, UK; 2013. p. 307-312. (In Eng.) doi: <https://doi.org/10.1109/IV.2013.39>
- [5] Bastian M., Heymann S., Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media*. 2009; 3(1):361-362. Available at: <https://ojs.aaai.org/index.php/ICWSM/article/view/13937> (accessed 17.09.2021). (In Eng.)
- [6] Ait El Mouden Z., Jakimi A., Hajar M. An Algorithm of Conversion Between Relational Data and Graph Schema. In: Rocha Á., Serrhini M. (eds.) *Information Systems and Technologies to Support Learning. EMENA-ISTL 2018. Smart Innovation, Systems and Technologies*. Vol. 111. Springer, Cham; 2019. p. 594-602. (In Eng.) doi: https://doi.org/10.1007/978-3-030-03577-8_65
- [7] Brandes U., Eiglsperger M., Lerner J., Pich C. Graph Markup Language (GraphML). In: Tamassia R. (ed.) *Handbook of Graph Drawing and Visualization*. 1st ed. Chapman and Hall/CRC, New York; 2013. p. 517-541. (In Eng.) doi: <https://doi.org/10.1201/b15385-19>
- [8] Borgman C.L. What are digital libraries? Competing visions. *Information Processing and Management*. 1999; 35(3):227-243. (In Eng.) doi: [https://doi.org/10.1016/S0306-4573\(98\)00059-4](https://doi.org/10.1016/S0306-4573(98)00059-4)
- [9] Fox E.A., Akscyn R.M., Furuta R.K., Leggett J.J. Digital libraries. *Communications of the ACM*. 1995; 38(4):22-28. (In Eng.) doi: <https://doi.org/10.1145/205323.205325>
- [10] Gonçalves M.A., Fox E.A., Watson L.T., Kipp N.A. Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. *ACM Transactions on Information Systems*. 2004; 22(2):270-312. (In Eng.) doi: <https://doi.org/10.1145/984321.984325>
- [11] Ferro N., Silvello G. NESTOR: A formal model for digital archives. *Information Processing & Management*. 2013; 49(6):1206-1240. (In Eng.) doi: <https://doi.org/10.1016/j.ipm.2013.05.001>



- [12] Constantin A., Peroni S., Pettifer S., Shotton D., Vitali F. The Document Components Ontology (Do-CO). *Semantic Web*. 2016; 7(2):167-181. (In Eng.) doi: <https://doi.org/10.3233/SW-150177>
- [13] Mann W.C., Thompson S.A. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text – Interdisciplinary Journal for the Study of Discourse*. 1988; 8(3):243-281. (In Eng.) doi: <https://doi.org/10.1515/text.1.1988.8.3.243>
- [14] Taboada M., Mann W.C. Rhetorical Structure Theory: Looking Back and Moving Ahead. *Discourse Studies*. 2006; 8(3):423-459. (In Eng.) doi: <https://doi.org/10.1177/1461445606061881>
- [15] Marcu D. The Rhetorical Parsing of Unrestricted Texts: A Surface-based Approach. *Computational Linguistics*. 2000; 26(3):395-448. (In Eng.) doi: <https://doi.org/10.1162/089120100561755>
- [16] Gagarin A.P., Serdyukov V.V. Dynamic ontology based retrieval and analysis of information on science and engineering. *Sovremennye informacionnye tehnologii i IT-obrazovanie = Modern Information Technologies and IT-Education*. 2018; 14(3):654-662. (In Russ., abstract in Eng.) doi: <https://doi.org/10.25559/SITITO.14.201803.654-662>
- [17] Fatimi S., El Saili C., Alaoui L. Semantic Oriented Text Clustering Based on RDF. *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*. IEEE Press, Fez, Morocco; 2020. p. 1-8. (In Eng.) doi: <https://doi.org/10.1109/ISCV49265.2020.9204133>
- [18] Khan J.A., Kumar S. Deep analysis for development of RDF, RDFS and OWL ontologies with protégé. *Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization*. IEEE Press, Noida, India; 2014. p. 1-6. (In Eng.) doi: <https://doi.org/10.1109/ICRITO.2014.7014747>
- [19] Brack A., Hoppe A., Stocker M., Auer S., Ewerth R. Requirements Analysis for an Open Research Knowledge Graph. *Proceedings of the 24th International Conference on Theory and Practice of Digital Libraries (TPDL 2020)*. Lyon, France; 2020. p. 1-15. (In Eng.) doi: <https://doi.org/10.48550/arXiv.2005.10334>
- [20] Färber M. The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data. In: Ghidini C., et al. (eds.) *The Semantic Web – ISWC 2019. ISWC 2019. Lecture Notes in Computer Science*. Vol. 11779. Springer, Cham; 2019. p. 113-129. (In Eng.) doi: https://doi.org/10.1007/978-3-030-30796-7_8
- [21] Brodaric B., Reitsma F., Qiang Y. SKling with DOLCE : toward an e-Science knowledge infrastructure. In: Eschenbach C., Gruninger M. (eds.) *Frontiers in Artificial Intelligence and Applications*. Vol. 193. Amsterdam, The Netherlands: IOS Press; 2008. p. 208-219. Available at: <https://biblio.ugent.be/publication/723935> (accessed 17.09.2021). (In Eng.)
- [22] Borlund P. Interactive Information Retrieval: An Introduction. *Journal of Information Science Theory and Practice*. 2013; 1(3):12-32. (In Eng.) doi: <https://doi.org/10.1633/JISaP.2013.1.3.2>
- [23] Shen J., Xiao J., He X., Shang J., Sinha S., Han J. Entity Set Search of Scientific Literature: An Unsupervised Ranking Approach. *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. ACM, New York, NY, USA; 2018. p. 565-574. (In Eng.) doi: <https://doi.org/10.1145/3209978.3210055>
- [24] Wu J., Williams K.M., Chen H.-H., Khabsa M., Caragea C., Tuarob S., Ororbis A.G., Jordan D., Mitra P., Giles C.L. CiteSeerX: AI in a Digital Library Search Engine. *AI Magazine*. 2015; 36(3):35-48. (In Eng.) doi: <https://doi.org/10.1609/aimag.v36i3.2601>
- [25] Otasek D., Morris J.H., Bouças J., Pico A.R., Demchak B. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biology*. 2019; 20:185. (In Eng.) doi: <https://doi.org/10.1186/s13059-019-1758-4>

Поступила 17.09.2021; одобрена после рецензирования 05.11.2021; принята к публикации 27.11.2021.

Submitted 17.09.2021; approved after reviewing 05.11.2021; accepted for publication 27.11.2021.

Об авторах:

Гагарин Андрей Петрович, профессор кафедры вычислительных машин, систем и сетей, Институт № 3 «Системы управления, информатика и электроэнергетика», ФГБОУ ВО «Московский авиационный институт (национальный исследовательский университет)» (125993, Российская Федерация, г. Москва, Волоколамское шоссе, д. 4), кандидат технических наук, профессор, **ORCID: <https://orcid.org/0000-0002-0929-2834>**, gagarin_ay@outlook.com

Филимонов Илья Андреевич, аспирант кафедры вычислительных машин, систем и сетей, Институт № 3 «Системы управления, информатика и электроэнергетика», ФГБОУ ВО «Московский авиационный институт (национальный исследовательский университет)» (125993, Российская Федерация, г. Москва, Волоколамское шоссе, д. 4), **ORCID: <https://orcid.org/0000-0002-2307-0453>**, ilafilimonov@mai.education

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the authors:

Andrey P. Gagarin, Professor of the Department of Computers, Systems and Networks, Institute of Control Systems and Computer Science in Engineering, Moscow Aviation Institute (National Research University) (4 Volokolamskoe shosse, Moscow 125993, Russian Federation), Cand.Sci. (Eng.), Professor, **ORCID: <https://orcid.org/0000-0002-0929-2834>**, gagarin_ay@outlook.com

Ilya A. Filimonov, Postgraduate Student of the Department of Computers, Systems and Networks, Institute of Control Systems and Computer Science in Engineering, Moscow Aviation Institute (National Research University) (4 Volokolamskoe shosse, Moscow 125993, Russian Federation), **ORCID: <https://orcid.org/0000-0002-2307-0453>**, ilafilimonov@mai.education

All authors have read and approved the final manuscript.

