

УДК 004.021, 004.42

ИНФОРМАЦИЯ О РОССИЙСКИХ НАУЧНЫХ ОРГАНИЗАЦИЯХ В МЕЖДУНАРОДНЫХ И РУССКОЯЗЫЧНЫХ ИСТОЧНИКАХ ДАННЫХ

З. В. Апанович^[0000-0002-5767-284X]

*Институт систем информатики им. А.П. Ершова Сибирского отделения
Российской академии наук, пр. Академика Лаврентьева, 6, Новосибирск,
Новосибирская обл., 630090*

apanovich_09@mail.ru

Аннотация

Рассмотрены международные и русскоязычные источники данных, предоставляющие информацию о российских научных организациях. Продемонстрировано, что русскоязычные источники данных содержат больше информации о русскоязычных научных организациях, чем англоязычные, но эта информация остается недоступной для англоязычных источников данных. Также описаны эксперименты по сопоставлению и интеграции информации о российских научных организациях в международных и российских источниках данных. Рассмотрены такие источники данных, как GRID, русскоязычная и англоязычная Wikipedia, Wikidata и eLIBRARY.ru. Работа является промежуточным этапом на пути к созданию открытого и расширяемого графа знаний.

Ключевые слова: *разноязычные графы знаний, идентификация сущностей, научные организации, корректность*

ВВЕДЕНИЕ

Современные программные средства позволяют разрабатывать не только гетерогенные графы знаний, но и графы знаний, собирающие информацию о сущностях определенного класса [1]. Центральный банк Италии создал граф знаний обо всех итальянских компаниях [2], граф знаний ClaimsKG [3] извлекает утверждения из веб-страниц, занимающихся проверкой фактов, таких как Politifact (<https://www.politifact.com/>), и публикует их, используя модель RDF, а EventKG [4] извлекает информацию о месте и времени всех сущностей, относящихся к классу Event (событие).

Информация о научных организациях является важным атрибутом, позволяющим идентифицировать авторов научных публикаций [5, 6], а также анализировать географическое распределение публикаций и оценивать влияние на цитируемость публикаций, связанную с географическим фактором [7]. К сожалению, информация о национальных научных организациях, например, о российских, часто является неполной или искаженной в международных базах данных. Поэтому весьма актуальной является разработка открытого и расширяемого графа знаний, посвященного российским научным организациям.

Одной из больших открытых англоязычных баз данных научных организаций является база данных GRID [8] (Global Research Identifier Database, <https://www.grid.ac/>). Она содержит информацию о более чем ста тысячах научных организациях по всему миру и используется не только как самостоятельный источник данных: данные GRID интегрированы в граф знаний SN SciGraph, разрабатываемый издательством Springer (<https://www.springernature.com/gp/researchers/scigraph>). Благодаря использованию данных GRID, все информация о публикациях SN SciGraph имеет географическую привязку (координаты, город, район, область, страна). Вторым достоинством GRID является наличие ссылок (поле External links) на глобальные библиографические ресурсы, такие как ROR (Research Organization Registry, <https://ror.org>), Crossref (<https://www.crossref.org/>), и ISNI (International Standard Name Identifier, <https://isni.oclc.org/>). Кроме ссылок на глобальные библиографические ресурсы, в поле External links имеются адрес веб-сайта организации и, опционально, ссылка на страницу организации в англоязычной Wikipedia. Поле Alternate Labels содержит варианты названия организации на английском языке и национальном языке конкретной организации. Приводятся как варианты полных названий (поле Aliases), так и сокращенных (поле Acronyms). Наконец, имеется возможность выстраивания иерархии организаций при помощи таких свойств, как дочерняя организация (child institutes) и вышестоящая организация (parent institute).

Помимо этого, каждый институт в GRID имеет метаданные, такие как уникальный идентификатор организации в GRID, тип организации (например, правительственная) и время возникновения организации (поле Established).

Список российских научных организаций, представленных в GRID, весьма ощутимо пополнился в 2021 году. Если в начале 2021 года он содержал данные о 2019 российских организаций, то уже в начале октября 2021 года там имелось 2099 организаций. К сожалению, информация о российских научных организациях, хранящаяся в GRID, с одной стороны, неполна, с другой стороны, содержит явные неточности.

Например, в GRID имеется страница, посвященная Сибирскому отделению Российской академии наук (Siberian Branch of the Russian Academy of Sciences, <https://www.grid.ac/institutes/grid.415877.8>). В качестве русскоязычного названия этой организации указан «Институт космофизических исследований и аэронауки им. Ю.Г. Шафера Сибирского отделения Российской академии наук», а в качестве дочерних организаций (“child institutes”) наряду с несколькими институтами, формально относящимися к СО РАН, перечислены образовательные организации различного подчинения, не имеющие к СО РАН никакого отношения. Среди них, например, имеется Восточно-сибирский институт Министерства внутренних дел Российской Федерации (East-Siberian Institute of the Ministry of Internal Affairs of the Russian Federation (<https://www.grid.ac/institutes/grid.445063.0>), Сибирский юридический институт ФСКН России (<https://www.grid.ac/institutes/grid.445537.4>) и др. Также можно заметить, что база данных GRID содержит в своем списке организации, которые уже прекратили свое существование. Например, у Федерального агентства научных организаций (Federal Agency for Scientific Organizations, <https://www.grid.ac/institutes/grid.484124.f>) указана дата возникновения (1913 год), но нет информации, что эта организация упразднена в 1918 году.

Наблюдая подобные неточности, можно предположить, что данная база пополнилась на основании англоязычных источников данных и, возможно, были бы полезными сопоставление и интеграция этих данных с русскоязычными источниками. Поскольку в GRID уже имеется возможность указания страницы организации в англоязычной Wikipedia, возникает предположение, что русскоязычная Wikipedia может содержать больше информации о российских научных организациях, и эта информация может быть лучшего качества, чем англоязычные источники данных.

Поэтому первый эксперимент состоял в установлении соответствия между страницами GRID, а также англоязычной и русскоязычной версиями Wikipedia при

помощи Wikipedia_API. Первоначально были проверены имеющиеся в GRID ссылки на англоязычную версию Wikipedia. Из 2019 российских организаций, имеющих в GRID, только 412 имели ссылки на страницы в англоязычной Wikipedia, и только 398 страниц англоязычной Wikipedia имели межязыковые ссылки (interlanguage links) на страницы русскоязычной Wikipedia. Для выяснения того, имеются ли в англоязычной и русскоязычной Wikipedia другие страницы русскоязычных организаций, описанных в GRID, осуществлялся поиск по таким атрибутам, как URL веб-сайта организации, название организации и др. К сожалению, поиск в Wikipedia по адресу веб-сайта организации оказался не очень эффективным, и в конечном итоге пришлось осуществлять поиск по различным версиям названия организации. Поиск осуществлялся как по англоязычному, так и по русскоязычному названию в обеих языковых версиях Wikipedia. В результате этого поиска было дополнительно обнаружено 674 страницы в русскоязычной Wikipedia, соответствующие организациям из GRID, из которых 353 страницы были связаны межязыковыми ссылками с англоязычной Wikipedia. В общей сложности было обнаружено 835 страниц Wikipedia, соответствующих русскоязычным организациям, представленным в GRID. Таким образом, этот эксперимент показал, что в русскоязычной версии Wikipedia действительно имеется больше информации о российских научных организациях, чем в англоязычной, но эта информация остается недоступной для англоязычных баз данных из-за их ориентированности на англоязычную версию Wikipedia.

Основная проблема состояла в том, что в русскоязычной Wikipedia оказалось мало явных ссылок на веб-сайты организаций, а поиск по названиям организаций сильно затруднен тем, что в разных базах данных приводятся разные названия одних и тех же организаций. Существующий, очень несовершенный алгоритм сопоставления планируется доработать.

Другим русскоязычным источником данных, заслуживающим внимания, является самая большая база данных российских научных организаций eLIBRARY.ru [9], на сайте которой представлено 12420 российских организаций.

1. СРАВНЕНИЕ ИНФОРМАЦИИ В GRID И eLIBRARY.RU

При сравнении информации в GRID и eLIBRARY.ru бросается в глаза, что количество организаций, представленных на сайте eLIBRARY.ru, в пять раз превышает количество организаций, представленных в GRID. Это соотношение остается постоянным для разных субъектов РФ. Так, если eLIBRARY.ru выдает список из 264 организации, расположенных в Новосибирске, то GRID показывает только 75 таких организаций, причем некоторые из представленных организаций относятся к разряду несуществующих.

На основе сопоставления количества данных о российских организациях, имеющихся в eLIBRARY.ru и GRID, возникает вопрос, за счет чего формируется такое значительное различие. При внимательном рассмотрении было обнаружено, что eLIBRARY.ru действительно содержит очень большой список российских организаций, но далеко не все организации из имеющегося списка относятся к научным. В список организаций входят все федеральные министерства и организации, подчиненные этим министерствам, региональные администрации, а также коммерческие организации, банки, больницы, индивидуальные предприниматели, заводы, комбинаты и прочие. Например, в eLIBRARY.ru можно обнаружить такую организацию, как «Домостроительный комбинат № 7», за которой не числится ни одной публикации и не имеется никакой информации, кроме почтового и юридического адресов. В целом в списке организаций eLIBRARY.ru примерно одна треть (4505 организаций) вообще не имеет никаких публикаций, а еще примерно тысяча организаций имеет по одной публикации и столько же – по две публикации.

В этой базе данных также имеется немало организаций, прекративших свое существование. Например, у многих институтов СО РАН в качестве ведомства, которому подчиняется данная организация, указано Федеральное агентство научных организаций (ФАНО). Страничка ФАНО имеется и в базе GRID, но в русскоязычной Wikipedia имеется информация об упразднении этой организации 15 мая 2018 года. Эта же информация продублирована в наборе данных Wikidata (Federal Agency for Scientific Organizations, Q16711297).

Заметим, что алгоритм сопоставления данных из eLIBRARY.ru и GRID на основе названий организаций обнаружил эквиваленты только для половины российских организаций, указанных в GRID. Предполагается, что такой невысокий

процент связан, прежде всего, с реальным различием в информации об организациях. В GRID, как и в eLIBRARY.ru, обнаружались организации, прекратившие свое существование, а также и заведомо ошибочные данные об организациях. С другой стороны, алгоритм сопоставления данных нуждается в улучшении. Наконец, возникает естественное желание использовать в качестве посредника между русскоязычными и англоязычными источниками данных ресурс, имеющий доступ к разноязычным источникам данных.

2. СОПОСТАВЛЕНИЕ ДАННЫХ eLIBRARY.RU И WIKIDATA

Примером очень перспективного источника данных является набор данных Wikidata. Wikidata возникла в 2014 году [10] как источник структурированных данных для управления фактами в различных языковых версиях Wikipedia. Разработчики планируют сделать ее центральной платформой управления Wikipedia, интегрируя данные из всех языковых «глав» Wikipedia. Интеграция осуществляется присваиванием не зависящего от конкретной языковой версии идентификатора каждому объекту реального мира и объединения всех высказываний о заданном объекте реального мира из всех языковых версий Wikipedia. Например, на странице Wikidata, посвященной СО РАН (<https://www.wikidata.org/wiki/Q3032414>), имеются ссылки на описание этой сущности в шести языковых версиях Wikipedia. Помимо интеграции информации из различных языковых версий Wikipedia этот набор данных так же, как и GRID, поддерживает ссылки на глобальные библиографические ресурсы при помощи указания идентификаторов в таких источниках данных, как VIAF (VIAF ID, свойство P214), идентификатор ISNI (ISNI ID, свойство P213), идентификатор Библиотеки Конгресса США (Library of Congress authority ID, свойство P244), идентификатор GRID (GRID ID, свойство P2427), идентификатор ROR (ROR ID, свойство P6782), основной государственный регистрационный номер (ОГРН, свойство P7011), идентификатор организации в eLIBRARY.ru (eLIBRARY.ru organisation ID, свойство P2463).

Также, помимо официального названия организации, в наборе имеются короткое название на русском и английском языках, информация о географическом положении (страна, область), даты возникновения и исчезновения сущности. Несмотря на эти явные достоинства, в наборе данных имеются и недостатки. Напри-

мер, на странице, посвященной СО РАН, приводится тот же самый неточный список подчиненных институтов, что и в базе данных GRID. В нем есть образовательные организации других ведомств, но зато не хватает некоторых институтов, реально подчиняющихся СО РАН. Например, в списке институтов, подчиненных СО РАН, нет ИСИ СО РАН, зато ИСИ СО РАН представлен как институт РАН, и на его странице (<https://www.wikidata.org/wiki/Q4201722>) почему-то указано, что институт назван не в честь академика А.П. Ершова, известного своими работами по информатике, а в честь Александры Петровны Ершовой (Alexandra Petrovna Ershova, Q60830445) – российского театрального педагога.

Следует отметить, что весьма незначительное количество российских организаций, представленных в наборе данных Wikidata, имеет на своих страницах идентификаторы из eLIBRARY.ru. Так, запуская запрос SPARQL относительно научных организаций (wd:Q16519632), расположенных на территории России и имеющих в wikidata идентификатор eLIBRARY.ru, мы получили список из восьмидесяти шести таких организаций, причем указаны, в основном, образовательные организации. Новосибирский государственный университет, например, имеет идентификатор 214. Пример запроса, выдающего научные организации с сайта Wikidata с идентификатором eLIBRARY.ru, показан на рисунке 1.

В то же время, среди 274 организаций, относящихся в наборе данных Wikidata к институтам РАН (wd:Q4201890), только четыре организации имеют идентификатор eLIBRARY.ru. Это означает, что задача сопоставления данных в вышеуказанных источниках является весьма актуальной. Далее будет описан наш подход к решению этой проблемы.

```
select ?s ?s_label ?elib
where {
  ?s wdt:P31/wdt:P279+ wd:Q16519632;
  wdt:P17 wd:Q159;
  wdt:P2463 ?elib;
  rdfs:label ?s_label filter(lang(?s_label) = 'ru').
}
```

Рис. 1. Запрос, выдающий научные организации, имеющие идентификатор eLIBRARY.ru на сайте wikidata.

Заметим, что неполнота данных, хранящихся в Wikidata, не позволяет в настоящий момент получить всю достоверную информацию при помощи запроса SPARQL. Так, например, запрос SPARQL, требующий выдать все российские научные организации, выдает только те организации, у которых в явном виде указано, что они расположены в России, а если сделать требование российской принадлежности необязательным (OPTIONAL), начинают выдаваться лишние организации, у которых вообще не указана страна принадлежности. Поэтому приходится программно решать задачу сопоставления сущностей, описанных в разных источниках данных.

На вход программы интеграции данных подается таблица, каждая строка которой соответствует одной организации в eLIBRARY.ru. В качестве атрибутов, извлекаемых из eLIBRARY.ru, берутся такие атрибуты, как полное название организации на русском языке, название на английском языке, русская аббревиатура названия, английская аббревиатура названия, страна, регион, русское название города, английское название города, почтовый адрес на русском языке, почтовый адрес на английском языке, юридический адрес, ведомство, тип организации, факс, URL сайта.

В результате работы алгоритма интеграции создается расширенная таблица, которая, кроме данных из eLIBRARY.ru, содержит для каждой организации информацию, удалось ли обнаружить ее в Wikidata. В случае, если организация обнаружена, данные об организации дополняются информацией, извлеченной из Wikidata. В частности, добавляются Wikidata_имя, Wikidata_идентификатор, Wikidata_алиасы, Wikidata_год основания, VIAF_Id, eLIBRARY_Id, GRID_Id. В качестве возможного результата также может появиться сообщение «организация не обнаружена» и «данных для идентификации недостаточно».

Общая схема алгоритма интеграции имеет вид:

1. Предварительная обработка названий организаций, имеющих в eLIBRARY.ru (перевод в нижний регистр, удаление стоп-слов «зао», «ооо», «им», циклическая замена некоторых слов при помощи словаря синонимов).

2. Поиск сущностей по одному из названий организации, указанных в eLIBRARY.ru (4 варианта названия, трансформированные названия, URL организации). В случае успешного поиска выдается файл JSON с краткой информацией о

найденном элементе, который позволяет извлечь дополнительную информацию о сущности: название, идентификатор, тип сущности, все доступные названия. В Wikidata эта информация содержится в поле «Также известный как».

3. Проверка, является ли сущность, найденная в Wikidata, эквивалентом организации из eLIBRARY.ru Для этого вычисляется коэффициент совпадения полных и трансформированных названий сущностей, выполняются сравнение URL организаций, проверка типа сущности, страны организации, местонахождения организации.

4. Пополнение исходной таблицы информацией из Wikidata

Сравнение названий организаций. При сравнении двух списков названий, извлеченных из eLIBRARY.ru и Wikidata, сначала осуществляется проверка на полное текстовое совпадение названий. В случае отрицательного результата проверки каждая строка-название разбивается на отдельные слова, они приводятся к нормальному виду при помощи модуля морфологического анализа. Для полученных строк вычисляется коэффициент совпадения по формуле:

$$Index = 2 \cdot (\text{количество совпадающих слов}) / (\text{суммарное количество слов}).$$

Вычисляется наилучший коэффициент совпадения по всем вариантам названий. Если этот коэффициент больше, чем 0,68 для некоторой пары названий, результат считается положительным, и осуществляется сравнение остальных атрибутов.

Сравнение URL организаций. При сравнении URL-адресов было замечено, что в базах данных он может быть записан по-разному. Например, в карточке организации *Администрация Архангельской области* в eLIBRARY.RU в качестве URL указан <http://www.dvinaland.ru/>, а в Wikidata – <https://dvinaland.ru/>. Таким образом, из-за различия конструкций “http” и “https” и наличия в одном из адресов конструкции “www” результат сравнения будет отрицательным. Во избежание таких ситуаций при сравнении URL указанные конструкции удаляются.

Проверка типа найденного объекта. Необходимо убедиться, что найденный объект является организацией. При помощи запроса SPARQL сгенерирован файл в формате CSV, содержащий названия всех подклассов класса Организация в онтологии Wikidata. Тип найденной сущности сравнивается с элементами этого файла. В случае отрицательного результата сравнения сущность отклоняется.

Проверка местонахождения организации. Если в качестве страны организации указана не Россия, поиск останавливается, в базу записывается, что организация находится не в России. Если Россия или другая страна не указаны на странице Wikidata, извлекается информация о местонахождении штаб-квартиры — в Wikidata чаще всего указывается город, также извлекается информация про административно-территориальное образование. Заметим, что в Wikidata вместо Москвы может быть указана, например, Московская область, что может привести к неправильному результату сравнения. Для решения этой проблемы используется информация об иерархии географических объектов. Был создан файл в формате JSON с распределением всех городов России по областям. Программа извлекает название города из Wikidata и, если оно не совпадает с положением, указанным в eLIBRARY.RU, то название города будет заменено на название области, в которой находится город. После этого географические положения сравниваются повторно. Если и в таком случае местоположения не совпадают, организация считается найденной неверно, и рассматривается следующая.

Таким образом, в настоящий момент организация считается верно идентифицированной в двух случаях:

А) Совпадают URL сайтов и некоторые пары вариантов названий.

Б) Сущность является организацией, названия организации в двух источниках совпадают, информация о сайтах не полна, организация находится в России. Для всех организаций, распознанных алгоритмом как идентичные, информация об этих организациях объединяется на основе расширения онтологии `schema.org`. На данный момент установлено соответствие между 3143 организациями Wikidata и eLibrary.ru. Получившийся экспериментальный источник данных будет в дальнейшем расширяться и интегрироваться с другими источниками данных, такими как GRID.

3. ЗАКЛЮЧЕНИЕ

Эксперименты с англоязычными и русскоязычными источниками данных показали, что русскоязычные источники информации содержат больше информации о русскоязычных научных организациях, чем англоязычные. К сожалению, эта информация остается в большинстве своем недоступной для англоязычных источ-

ников данных. Для решения этой проблемы разработан метод интеграции информации из разноязычных источников данных. Создана экспериментальная версия базы данных научных организаций, состоящая из 3143 научных организаций. Планируется превратить эту базу в открытый и расширяемый граф знаний. Авторы также считают, что для поддержания полноты и корректности информации о научных организациях каждая научная организация должна поддерживать на международных платформах свою страницу с указанием всех идентификаторов организации.

СПИСОК ЛИТЕРАТУРЫ

1. *Апанович З.В.* Эволюция понятия и жизненного цикла графов знаний // Системная информатика. 2020. № 16. С. 57–74.
 2. *Atzeni P., Bellomarini L., Iezzi M., Sallinger E., Vlad A.* Weaving Enterprise Knowledge Graphs: The Case of Company Ownership Graphs. URL: https://openproceedings.org/2020/conf/edbt/paper_334.pdf
 3. *Tchechmedjiev A., Fafalios P., Boland K., Gasquet M., Zloch M., Zapilko B., Dietze S., Todorov K.* ClaimsKG: A knowledge graph of fact-checked claims. In International Semantic Web Conference, Springer, 2019. P. 309–324.
 4. *Gottschalk S., Demidova E.* EventKG: A Multilingual Event-Centric Temporal Knowledge Graph. URL: <https://arxiv.org/pdf/1804.04526.pdf>
 5. *Apanovich Z.* Matching of authors and publications in multilingual bibliographic knowledge bases // CEUR Workshop Proceedings. SSI 2019. Proceedings of the 21st Conference on Scientific Services and Internet. 2020. P. 26–37.
 6. *Hajra A., Radevski V., Tochtermann K.* Author profile Enrichment for Cross-linking Digital Libraries // Research and Advanced Technology for Digital Libraries Springer International Publishing. Lecture Notes in Computer Science. 2015. Vol. 9316. P. 124–136.
 7. *Mannocci A., Osborne F., Motta E.* Geographical trends in academic conferences: An analysis of authors' affiliations // Data Science. 2019. 2(1) P. 181–203.
 8. Global Research Identifier Database. URL: <https://www.grid.ac/>
 9. Научная электронная библиотека eLIBRARY.ru.
URL: <https://www.elibrary.ru/>
-

10. *Ismayilov A., Kontokostas D., Auer S., Lehmann J., Hellmann S.* Wikidata through the Eyes of DBpedia.

URL: <http://www.semantic-web-journal.net/system/files/swj1462.pdf>

11. *Lenzerini M.* Data Integration: A Theoretical Perspective Conference: Proceedings of the Twenty-first ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems. 2002. P. 233–246.

INFORMATION ABOUT RUSSIAN RESEARCH ORGANIZATIONS IN MULTILINGUAL DATA SOURCES

Z. V. Apanovich^[0000-0002-5767-284X]

A.P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences, Novosibirsk State University, Novosibirsk

apanovich_09@mail.ru

Abstract

International and Russian-language data sources that provide information about Russian research-related organizations are considered. It is demonstrated that Russian-language data sources contain more information about Russian research-related organizations than most international data sources, but this information remains unavailable for English-language data sources. Experiments on comparison and integration of information about Russian research organizations in international and Russian data sources are outlined. Data sources such as GRID, Russian and English chapters of Wikipedia, Wikidata and eLIBRARY.ru are considered. The work is an intermediate step towards the creation of an open and extensible knowledge graph.

Keywords: *multi-lingual knowledge graphs, identity resolution, research-related organizations, correctness*

REFERENCES

1. *Apanovich Z.V.* Evolyuciya ponyatiya i zhiznennogo cikla grafov znanij // Sistemnaya informatika. 2020. № 16. S. 57–74.

2. *Atzeni P., Bellomarini L., Iezzi M., Sallinger E., Vlad A.* Weaving Enterprise Knowledge Graphs: The Case of Company Ownership Graphs. URL: https://openproceedings.org/2020/conf/edbt/paper_334.pdf

3. *Tchechmedjiev A., Fafalios P., Boland K., Gasquet M., Zloch M., Zapilko B., Dietze S., Todorov K.* ClaimsKG: A knowledge graph of fact-checked claims. In *International Semantic Web Conference*, Springer, 2019. P. 309–324.

4. *Gottschalk S., Demidova E.* EventKG: A Multilingual Event-Centric Temporal Knowledge Graph. URL: <https://arxiv.org/pdf/1804.04526.pdf>

5. *Apanovich Z.* Matching of authors and publications in multilingual bibliographic knowledge bases // *CEUR Workshop Proceedings. SSI 2019 – Proceedings of the 21st Conference on Scientific Services and Internet*. 2020. P. 26–37.

6. *Hajra A., Radevski V., Tochtermann K.* Author profile Enrichment for Cross-linking Digital Libraries // *Research and Advanced Technology for Digital Libraries Springer International Publishing. Lecture Notes in Computer Science*. 2015. Vol. 9316. P. 124–136.

7. *Mannocci A., Osborne F., Motta E.* Geographical trends in academic conferences: An analysis of authors' affiliations // *Data Science*. 2019. 2(1) P. 181–203.

8. Global Research Identifier Database. URL: <https://www.grid.ac/>

9. eLIBRARY.ru. URL: <https://www.elibrary.ru/>

10. *Ismayilov A., Kontokostas D., Auer S., Lehmann J., Hellmann S.* Wikidata through the Eyes of DBpedia.

URL: <http://www.semantic-web-journal.net/system/files/swj1462.pdf>

11. *Lenzerini M.* Data Integration: A Theoretical Perspective Conference: Proceedings of the Twenty-first ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems. 2002. P. 233–246.

СВЕДЕНИЯ ОБ АВТОРЕ



АПАНОВИЧ Зинаида Владимировна – старший научный сотрудник Института систем информатики СО РАН, доцент Новосибирского государственного университета. Сфера научных интересов – визуализация информации, визуализация графов, Semantic Web.

Zinaida Vladimirovna APANOVICH – senior researcher of the Institute of Informatics Systems of SB RAS, Associate Professor of Novosibirsk State University. Research interests include information visualization, graph visualization, Semantic Web.

email: apanovich@iis.nsk.su

ORCID 0000-0002-5767-284X

Материал поступил в редакцию 25 октября 2021 года