

Организация поиска в базе данных со связанными сущностями

К.П. Погорелко^{1, 2}✉, Г.И. Савин^{1, 2}

¹ Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия

² Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

Ссылка для цитирования

Погорелко К.П., Савин Г.И. Организация поиска в базе данных со связанными сущностями // Программные продукты и системы. 2024. Т. 37. № 4. С. 524–531. doi: 10.15827/0236-235X.148.524-531

Информация о статье

Группа специальностей ВАК: 2.3.5

Поступила в редакцию: 24.07.2024

После доработки: 22.08.2024

Принята к публикации: 30.08.2024

Аннотация. В базе данных со связанными сущностями возникает проблема организации поиска, при котором можно не только формулировать условия к атрибутам определенной сущности, но и задавать условия на связанные с ней сущности. В статье описан реализованный поисковый механизм, на базе которого организован поиск в информационной системе электронной библиотеки «Научное наследие России». Поисковый механизм обеспечивает работу на русском и английском языках. Настройка поискового механизма на структуру конкретной базы данных заключается в заполнении ряда таблиц. Поисковый механизм позволяет работать с такими типами атрибутов, как текстовое поле, числовое значение, булево значение, полнотекстовый индекс текстового поля, кодификаторы. Предусмотрен различный функционал для администраторов и обычных пользователей. Для формулировки запроса к базе данных разработан интуитивно понятный интерактивный построитель поисковых запросов, который позволяет в режиме диалога задавать условия к первоначально выбранной сущности, к связанным с ней сущностям, к их количеству и типам связи. Для работы с кодификаторами большого объема предусмотрен расширенный функционал. Описана организация html-разметки формы ввода и структуры данных, формируемых при построении запроса. Предложено алгоритмическое решение, которое позволяет построить SQL-предписание к базе данных по сформулированному в построителе запросу. Приведены пример запроса к связанным сущностям и построенное для этого запроса SQL-предписание. Описана организация системы визуализации результатов поиска. Предложенные алгоритмы построения поискового механизма могут служить моделью для организации поиска в Едином цифровом пространстве научных знаний, включающем разнородные именованные связи между объектами.

Ключевые слова: информационный поиск, документный поиск, база данных, автоматизированная система, связанные сущности, графовые системы

Благодарности. Работа выполнена в МСЦ РАН и НИЦ «Курчатовский институт» в рамках госзадания № FNEF-2024-0014

Введение. Начало проекту российской академической научной *электронной библиотеки «Научное наследие России»* (ЭБ ННР) было положено в 2009 г. [1]. Он создавался как интегратор многоаспектной цифровой информации о выдающихся российских ученых, внесших вклад в развитие фундаментальных естественных и гуманитарных наук [2]. По сравнению с традиционными цифровыми библиотеками БД этой системы имеет расширенный набор сущностей, в нее добавлены такие, как «музейный объект» и «коллекция». Кроме того, традиционные сущности имеют расширенный набор атрибутов, например, сущность «персона» имеет атрибуты, отражающие биографические сведения, направления научной деятельности и т.п. [3]. Первоначально поиск в ЭБ ННР был организован средствами программной оболочки Единого научного информационного пространства РАН [4] с достаточно ограниченными возможностями. Поскольку предполагается, что структура БД ЭБ ННР будет расширяться как модель Единого цифрового пространства научных знаний [5], требуется реализация по-

искового механизма, предоставляющего больше возможностей при формулировке поискового запроса.

В информационных системах, БД которых содержат связанные сущности, возникает потребность организации такого поиска, при котором обеспечивается возможность формулировать условия к атрибутам определенной сущности, а также задавать условия на количество связанных с ней сущностей, отвечающих своему набору заданных условий. Эти условия, в свою очередь, могут содержать требования на связи с какими-то другими сущностями. Кроме того, связи между сущностями могут иметь спецификации. Например, спецификация связи между сущностями «публикация» и «персона» может принимать значения «автор», «редактор», «переводчик», «о нем» и т.п. Таким образом, необходим поисковый механизм, который позволял бы формулировать подобный следующему запрос: «получить список авторов, публикации которых в области физики входят в коллекцию «Материалы по истории Российской академии наук».

В настоящее время для организации информационного поиска используется ряд моделей. Лидирующее положение среди них занимают векторная модель, вероятностные модели, такие как LSI и LDA, и предварительно обученные модели на основе искусственного интеллекта, например, BERT [6]. Все эти модели предназначены для организации поиска в слабо или совсем не структурированной текстовой информации большого объема. Также отмечается их основной недостаток – невозможность четко сформулировать требования к получаемым документам, что является актуальным, особенно в академической среде [7]. В силу этого для организации поиска для ЭБ ННП была выбрана булева модель. Сформулированный ранее запрос будет выглядеть в ней так: «найти экземпляры сущности «персона», которые имеют связь, специфицированную как «автор», хотя бы с одним экземпляром сущности «публикация», которая, в свою очередь, во-первых, отнесена к проблематике «физика» и, во-вторых, имеет связь хотя бы с одним экземпляром сущности «коллекция», название которого соответствует требуемому».

Известно, что БД, обеспечивающие работу с многосвязными данными, относят к категории графовых [8]. Поскольку использование стандартного языка SQL при формировании запросов к связанным сущностям приводит к достаточно громоздким конструкциям, для упрощения манипулирования данными в графовых базах разработан ряд специализированных языков. Прежде всего необходимо отметить CQL (*Cypher Query Language*), созданный в рамках проекта Neo4j (<https://technology.amis.nl/database/querying-connected-data-in-graph-databases-with-neo4j>), и GQL (*Graph Query Language*) (<https://github.com/OlofMorra/GQL-parser/blob/main/src/main/resources/report/A%20Semantics%20of%20GQL;%20a%20New%20Query%20Language%20for%20Property%20Graphs%20Formalized.pdf>), который в апреле 2024 г. был опубликован ISO в качестве международного стандарта. Однако использование этих языков требует их изучения, знания специфики и структуры данных конкретной базы и предоставления соответствующих полномочий пользователю. Задачей описываемой разработки было создание поискового механизма с интуитивно понятным интерфейсом, который позволял бы формулировать запросы неподготовленному пользователю.

Организация поиска потребовала решения трех задач. Во-первых, необходим собственно

построитель поисковых запросов, который, предоставляя на каждом этапе выбор из возможных для этого места опций, позволяет пользователю формулировать требуемый запрос. Во-вторых, требуется алгоритмическое решение для преобразования сформулированного в построителе запроса в SQL-предписание к БД. В-третьих, необходима система визуализации найденных результатов.

Вариант поисковой системы с такими возможностями был реализован в ЭБ ННП [9]. С учетом опыта эксплуатации библиотеки в настоящее время появился новый вариант поисковой системы, отличающийся расширенным функционалом и возможностями его быстрой перенастройки на изменения структуры БД. Построитель поисковых запросов и система выдачи результатов реализованы на стороне клиента по технологии одностраничного приложения (SPA) на языке JavaScript. Сервисы для них на стороне сервера и функционал построения поискового предписания для SQL-сервера реализованы на языке C# в среде .NET 8.0. БД располагается на сервере PostgreSQL. Поисковая система для пользователей доступна по адресу <http://www.e-heritage.ru/search/>.

Структура данных

Настройка поискового механизма на структуру конкретной БД сводится к заполнению ряда таблиц.

Прежде всего это перечень сущностей БД (классов объектов в терминологии онтологии Единого цифрового пространства научных знаний [10]). Для каждой сущности, которая может участвовать в поиске, задается следующая информация.

- Название таблицы БД, содержащей экземпляры сущности. Это название является идентификатором сущности, на который ссылаются другие таблицы настройки поискового механизма.

- Название поля таблицы, являющегося главным индексом.

- Название сущности на двух языках. Это связано с тем, что ЭБ ННП является двуязычной и построитель запросов работает в зависимости от затребованного языка.

- Признак, указывающий на то, что данная сущность доступна только администраторам системы. БД содержит данные, необходимые для обеспечения технологического процесса по вводу/редактированию информации, поэтому

в ней есть, например, такие сущности, как «ошибка», которые обычным пользователям не должны быть доступны.

– Условие на атрибуты, что данный экземпляр сущности имеет статус «опубликовано» и может быть выдан обычному пользователю. Администратору выдаются все найденные записи.

– Информация, необходимая для представления пользователю найденных результатов. Она включает перечень полей, из которых выбираются сведения, строку форматирования этих полей и форматы http-ссылок на конкретную сущность для администраторов и обычных пользователей.

– Перечень возможных сортировок сущностей с указанием названия сортировки на двух языках и перечня полей, по которым производится сортировка полученных результатов поиска.

– Перечень возможных элементов поискового условия для данной сущности.

Определяется перечень кодификаторов системы, которые могут использоваться при формулировке поискового запроса. Для каждого кодификатора задаются названия

– таблицы, содержащей кодификатор; это название является идентификатором кодификатора;

– кодификатора на русском и английском языках;

– полей, содержащих русские и английские тексты кодификатора;

– индексного поля.

Задается перечень связей между сущностями. Для каждой связи приводится следующая информация:

– название таблицы, связывающей сущности (оно является идентификатором связи);

– идентификаторы связанных сущностей;

– названия индексных полей, по которым осуществляется связь;

– если связь может иметь спецификации, указывается идентификатор соответствующего кодификатора.

Для каждого возможного элемента поискового условия для сущности задается следующая информация.

– Идентификатор элемента.

– Название на русском и английском языках.

– Признак того, что данный элемент доступен только администраторам.

– Тип элемента. Элемент может быть условием на значения как какого-либо атрибута

экземпляра сущности, так и связанных с ней кодификаторов. Он также может быть условием на связанную сущность, а точнее, на количество связанных с этим экземпляром других экземпляров связанных сущностей, отвечающих своему набору условий. Поддерживаются следующие типы элементов: текстовый, числовой, полнотекстовый индекс текстового поля, дата, двоичный (возможные значения «Да»–«Нет»), простой кодификатор – кодификатор небольшого размера, который может быть показан целиком, иерархический кодификатор – кодификатор большого размера с иерархической структурой, как, например, кодификатор ГРНТИ, условие на связанную сущность.

– Дополнительная информация. Для элемента, относящегося к условию на атрибут сущности, задается перечень названий полей, в которых содержится соответствующая информация. В определенных случаях поиск целесообразно выполнять сразу по нескольким полям. Так, в ЭБ ННР название публикации отражено в полях «оригинальное название» и «переведенное название», название коллекции присутствует на двух языках и т.д. Таким образом, условие, заданное для элемента «название публикации», проверяется как для поля «оригинальное название», так и для поля «переведенное название». Для кодификатора задаются идентификатор кодификатора и информация о таблице и индексах, связывающих кодификатор с сущностью. Для условия на связанную сущность указывается идентификатор связи.

Вся работа функциональных блоков поисковой системы основана на использовании данной информации.

Построитель поискового запроса

При запросе на поиск построитель поискового запроса запрашивает у соответствующего сервиса на сервере поисковую информацию, передавая в качестве параметра код языка, на котором общается пользователь. На основе этого параметра, а также аутентификации пользователя как администратора сервис возвращает выборку из структуры данных, соответствующую запрошенным параметрам. Далее с учетом этой информации в зависимости от действий пользователя происходит модификация html-кода страницы и формируются необходимые структуры данных. По умолчанию сущностью для поиска выбирается первая из списка сущностей, а условие соответствует первому элементу из списка возможных элементов

поискового условия. При выборе другой сущности для поиска автоматически изменяется и условие по умолчанию.

При нажатии кнопки «Добавить условие...» дополнительные условия располагаются построчно с тем же отступом и имеют возможность выбора связывающего оператора И/ИЛИ. Перед каждым условием возможен выбор оператора НЕ. Вычисление логического значения ведется сверху вниз без учета старшинства операторов. Для возможности редактирования запроса каждая поисковая строка имеет кнопку, позволяющую удалить эту строку. В зависимости от типа выбранного элемента поискового условия формируются соответствующие элементы html-разметки.

Набор элементов для условий на атрибуты сущности:

- для текстового атрибута – возможность выбора опции «содержит», «начинается с» или «равно» и поля для ввода необходимых фрагментов текста;

- для числового атрибута – возможность выбора оператора «=», «>» или «<» и поля для ввода числового значения; поскольку числовые поля не всегда могут быть определены (например, если для года введено «первая половина XVIII века»), для обеспечения полноты поиска вводится возможность добавления к результату записи с неустановленным значением;

- для полнотекстового индекса – поле, позволяющее ввести слова, словоформы которых будут обнаруживаться в индексе, и возможность выбора взаимного расположения словоформ «рядом» или «в любом порядке»;

- для атрибута типа дата – возможность выбора оператора «=», «>» или «<» и поля для задания даты;

- для двоичного типа – возможность выбора «да»–«нет».

Для выбора условий на значения кодификатора имеется кнопка «Выбрать». Дальнейшие операции при ее нажатии зависят от типа кодификатора.

- Для простого кодификатора он выдается целиком. После завершения выбора в поисковой строке отображаются выбранные тексты рубрик. Повторное нажатие кнопки позволяет редактировать сделанный выбор.

- Для иерархического кодификатора предусмотрена более сложная функциональность, позволяющая производить поиск как по иерархии, так и по текстам рубрик. При выборе рубрики, содержащей подчиненные рубрики, можно указать необходимость их включения в условие. Поскольку выбранных рубрик может быть много, по завершении выбора в поисковой строке выдается только их количество. Повторное нажатие кнопки позволяет редактировать сделанный выбор.

Пример задания поисковых условий разного типа приведен на рисунке.

На связанную сущность имеются две группы условий: на количество связанных с этим экземпляром сущностей, отвечающих определенным условиям, и собственно эти условия. При выборе в поисковой строке элементов формируются соответствующие поля.

Условие на количество может иметь значения «хотя бы одна», «все» или количественное

Электронная библиотека "Научное наследие России"

Главная Поиск Выставки Фотоархив ИЭА РАН Контакты О проекте

Поиск

Найти Публикации С условием

Название содержит []

И Год издания [] Добавить документы с неустановленным годом []

И Оглавление содержит слова [] Расстояние между словами любое []

И Язык публикации [] латинский, русский Выбрать []

И Научная тематика [] Выбрано рубрик: 1 Выбрать []

Добавить условие для Публикации

Найти Сортировать по [] возрастанию [] Документов на странице 10 [] Очистить форму

© 2020-2023 - МСЦ РАН

Использование поисковых условий разного типа

Usage of search terms of different types

значение с указанием числа и возможности выбора оператора « \Leftarrow », « \Rightarrow » или « \Leftarrow ». Если связь между сущностями может иметь спецификацию, следует нажать кнопку для ее выбора. В результате выдается список спецификаций данной связи из соответствующего кодификатора, после выбора нужных значений их названия выводятся в соответствующую поисковую строку.

Набор условий на связанную сущность содержит поисковое условие по умолчанию для данной сущности и кнопку добавления условий. Этот список форматируется с отступом по отношению к родительскому списку условий, что позволяет наглядно представлять структуру формируемого запроса.

Возможности системы проиллюстрированы дополнительными примерами запросов (<http://www.swsys.ru/uploaded/image/2024-4/11.jpg>, <http://www.swsys.ru/uploaded/image/2024-4/12.jpg>).

Структура информации строителя поискового запроса

Информация, вводимая пользователем в ходе конструирования поискового запроса, формируется следующим образом. Каждая поисковая строка получает уникальный номер. Все поля и элементы, относящиеся к данной строке, получают при генерации идентификатор (атрибуты 'id' и 'name' соответствующего html-элемента), состоящий из префикса, который определяет тип поля или элемента, и из номера поисковой строки. Так, для элемента, определяющего выбранный поисковый элемент, используется префикс 'f'. Значения его атрибута 'value' содержит идентификатор элемента в перечне возможных поисковых элементов сущности. Для поля, содержащего введенные данные, используется префикс 'd', для элемента, определяющего отношение данных, – префикс 't', для элемента связи условий – 'a', для отрицания – 'n' и т.д. Структура формируемого запроса сохраняется в виде дерева, каждый элемент которого соответствует строке поискового условия и содержит информацию о собственном числовом идентификаторе, а если это условие на связанную сущность, то еще и список идентификаторов строк, задающих условия на эту сущность. При запросе на поиск на сервер передаются структура запроса, свернутая в формат JSON, и вся информация, введенная пользователем в параметрах запроса в виде словаря «ключ–значение», где ключом является идентификатор соответствующих html-элементов, а значением – значение их атрибута 'value'.

Генератор поискового предписания

Генерация поискового предписания для SQL-сервера происходит следующим образом. Определяется сущность, экземпляры которой требуется найти. Далее вызывается рекурсивная процедура построения SQL-запроса, которой в качестве параметров передаются идентификатор сущности и соответствующий ей узел структуры запроса.

Представим алгоритм процедуры. По идентификатору сущности на основе информации из перечня сущностей строится оператор SQL: 'SELECT <главный индекс таблицы сущности> FROM <таблица сущности> WHERE'. Затем перебираются номера дочерних элементов узла структуры запроса. Для каждого номера n из присланного в запросе словаря выбирается элемент с ключом 'f<n>', на основе которого отбирается информация из списка поисковых элементов соответствующей сущности. Далее в зависимости от типа элемента формируется поисковое условие.

- Для элементов, имеющих тип «текст», анализируется заданный тип отношения (элемент 't<n>'). Если задано отношение «содержит», то введенные данные (элемент 'd<n>') разбиваются на подстроки и для каждой подстроки формируется поисковое условие, все они затем соединяются оператором AND. Для типов «равно» и «начинается с» данные используются непосредственно. Для поиска в базе используется SQL-оператор 'ilike', который обеспечивает сравнение без учета регистра с заданием усечения, соответствующего указанному типу отношения. Для административного варианта поиска проверяется дополнительное условие. Если задано отношение «равно», а поле данных пустое, то формируется запрос на документы, у которых это поле не заполнено.

- Для элементов, имеющих типы «число», «дата» или «булево», проверяется корректность информации, заданной в соответствующем поле, и формируется соответствующее поисковое условие. В случае типа «число», если задана опция добавления документов с неопределенным значением, добавляется соответствующее условие.

- Для кодификаторов на основе информации об отобранных в запросе индексах (элемент 'd<n>') и о таблице связи кодификатора и сущности формируется условие '<главный индекс таблицы сущности> IN (SELECT <главный индекс таблицы сущности > FROM

<таблица связи с кодификатором> WHERE <индекс кодификатора> IN (<перечень отобранных индексов>))’.

- Для связанной сущности выбирается ее идентификатор и осуществляется рекурсивный вызов процедуры построения SQL-запроса с этим идентификатором и соответствующим элементом поисковой структуры. Возвращенное значение (оператор SELECT для связанной сущности) дополняется необходимыми параметрами, соответствующими количественным условиям на связанные сущности и таблицы связи сущностей.

Если в поисковой строке задан оператор НЕ (элемент ‘n<n>’), то полученное поисковое условие предваряется оператором NOT и на основе элемента ‘a<n>’ добавляется к генерируемому WHERE с использованием оператора AND или OR.

Если поисковый запрос не от администратора, после обработки всех дочерних элементов добавляется условие, что данный экземпляр опубликован, и процедура возвращает сформированный SQL-оператор.

Возвращенная процедурой команда для главной сущности дополняется параметрами выбранной пользователем сортировки и передается SQL-серверу. Полученные в результате выполнения запроса индексы передаются системе выдачи результатов поиска.

Сгенерированная SQL-команда для запроса на связанные сущности (<http://www.swsys.ru/uploaded/image/2024-4/7.jpg>) выглядит следующим образом:

```
SELECT personid FROM person WHERE
(person.personid in (SELECT personid
FROM publicationperson WHERE (publicationperson.roleid in (1)) AND (publicationid in (SELECT publicationid FROM
publication WHERE (((publication.publicationid in (SELECT publicationid FROM
publicationcollection WHERE collectionid in (SELECT collectionid FROM collection WHERE ((collectionname ilike '%академ%') and (collectionname ilike '%наук%')) OR ((collectionnameeng ilike '%академ%') and (collectionnameeng ilike '%наук%')))) and (ispublished)))) AND (publication.publicationid in (select distinct publicationid from publicationgrnti where codgrnti like '29%')))) and (ispublished)))) and (ispublished)
```

Система визуализации результатов поиска

Поскольку объем полученной после выполнения запроса информации может быть значительным, в системе организована страничная выдача. К ее организации обычно используется подход, когда, при каждом запросе новой страницы выполняется процедура поиска и из нее выделяется нужный диапазон записей. Поскольку БД в ЭБ ННР меняется динамически, данный подход приводит к неприятным последствиям, так как результаты одного и того же запроса, полученные в разные моменты времени, могут отличаться и одна и та же страница при повторной выдаче может содержать разные документы. Кроме того, запросы бывают достаточно сложными и их повторное вычисление при пролистывании может привести к неоправданным задержкам. В данной поисковой системе результат поиска в виде списка найденных идентификаторов передается клиенту. Клиент при запросе страницы пользователем выделяет нужный диапазон идентификаторов и запрашивает у сервера необходимые документы. Сервер, получив нужные идентификаторы, на основе информации из описания сущности о формате представления пользователю найденных результатов и вида запроса (администратор–пользователь) формирует ответ, который возвращает клиенту.

Выводы

Анализ протоколов работы поискового механизма позволяет сделать вывод, что при наличии ссылок на прежние, более привычные формы он пользуется устойчивым спросом, хотя и скрыт за ссылкой «расширенный поиск». Основными потребителями являются пользователи, у которых есть необходимость в получении ответов на сложные запросы, – администраторы системы и специалисты, занимающиеся аналитикой.

Разработанные подходы к формированию поисковых механизмов планируется реализовать в Едином цифровом пространстве научных знаний.

Список литературы

1. Каленов Н.Е., Савин Г.И., Сотников А.Н. Электронная библиотека «Научное наследие России» // Информационные ресурсы России. 2009. № 2. С. 19–20.
2. Каленов Н.Е., Савин Г.И., Сотников А.Н. Электронная библиотека «Научное наследие России» как интегратор научной информации // Информационные системы и процессы: сб. науч. тр. 2016. Т. 15. С. 21–29.

3. Кириллов С.А., Соболевская И.Н., Сотников А.Н. Принципы формирования и представления междисциплинарных коллекций в цифровом пространстве научных знаний // Электронные библиотеки. 2021. Т. 24. № 2. С. 294–314. doi: 10.26907/1562-5419-2021-24-2-293-313.
4. Бездушный А.А., Бездушный А.Н., Серебряков В.А., Филиппов В.И. Интеграция метаданных Единого научного информационного пространства РАН. М.: изд-во ВЦ РАН, 2005. 238 с.
5. Каленов Н.Е., Погорелко К.П., Сотников А.Н. О развитии электронной библиотеки «Научное наследие России» как составляющей Единого цифрового пространства научных знаний // Информационные процессы. 2022. Т. 22. № 3. С. 155–166.
6. Lin J., Nogueira R., Yates A. Pretrained Transformers for Text Ranking: Bert and Beyond. Springer Cham Publ., 2022, 307 p. doi: 10.1007/978-3-031-02181-7.
7. Хьерланд Б. Информационный поиск и организация знаний: взгляд с точки зрения философии науки // Междунар. форум по информации. 2022. Т. 47. № 4. С. 36–57. doi: 10.36535/0203-6460-2022-04-3.
8. Robinson I., Webber J., Eifrem E. Graph Databases. O'Reilly Publ., CA, 2015, 220 p.
9. Погорелко К.П. Новая версия программного обеспечения электронной библиотеки «Научное наследие России» // Информационные ресурсы России. 2020. № 5. С. 27–29.
10. Каленов Н.Е., Сотников А.Н. Структура онтологии единого цифрового пространства научных знаний // Научно-техническая информация. Сер. 2. 2023. № 7. С. 20–26. doi: 10.36535/0548-0027-2023-07-3.

Software & Systems

doi: 10.15827/0236-235X.148.524-531

2024, 37(4), pp. 524–531

Search organization in a database with related entities

Konstantin P. Pogorelko^{1,2}✉, **Gennady I. Savin**^{1,2}¹ Joint Supercomputer Center of RAS, Moscow, 119334, Russian Federation² National Research Centre “Kurchatov Institute”,
Moscow, 123182, Russian Federation**For citation**Pogorelko, K.P., Savin, G.I. (2024) ‘Search organization in a database with related entities’, *Software & Systems*, 37(4), pp. 524–531 (in Russ.). doi: 10.15827/0236-235X.148.524-531**Article info**

Received: 24.07.2024

After revision: 22.08.2024

Accepted: 30.08.2024

Abstract. There is the problem of search organization in a database with related entities, in which it is possible to formulate conditions not only to the attributes of a certain entity, but also to set conditions on related entities. The paper describes the realized search mechanism. On the basis of this mechanism the author organized the search in the information system of the digital library “Scientific Heritage of Russia”. The search mechanism provides the work in the Russian and English languages. The author customize the search engine to the structure of a particular database by filling in a number of tables. The search engine allows working with such types of attributes as text field, numeric value, boolean value, full-text index of text field, codifiers. There are various functionalities for administrators and regular users. To formulate a query to the database, the author designed an intuitive interactive search query builder. It allows you to set conditions for the initially selected entity, related entities, their number and types of connection in the dialog mode. There is an extended functionality for working with codifiers of large volume. The author described the html markup organization of the input form and the data structure formed during query building. He gave an algorithmic solution that allows building an SQL instruction to the database according to the query formulated in the builder. He also gave an example of a query to related entities and the SQL instruction built for this query. The author described the organization of the search results visualization system. The proposed algorithms for building a search mechanism can serve as a model for organizing a search in the Unified Digital Space of Scientific Knowledge, including heterogeneous named relationships between objects.

Keywords: information search, document search, database, automated system, related entities, graph systems**Acknowledgements.** The paper was carried out under the government assignment, project no. FNEF-2024-0014**References**

1. Kalenov, N.E., Savin, G.I., Sotnikov, A.N. (2009) “‘Russian scientific heritage’ digital library as an integrator of scientific information’, *Information Resources of Russia*, (2), pp. 19–20 (in Russ.).
2. Kalenov, N.E., Savin, G.I., Sotnikov, A.N. (2016) “‘Russian scientific heritage’ digital library as an integrator of scientific information’, *Proc. Information Systems and Processes*, 15, pp. 21–29 (in Russ.).

3. Kirillov, S.A., Sobolevskaya, I.N., Sotnikov, A.N. (2021) 'Some aspects of the formation and representation principle of interdisciplinary collection in the digital space of scientific knowledge', *Russian Digital Libraries J.*, 24(2), pp. 294–314 (in Russ.). doi: 10.26907/1562-5419-2021-24-2-293-313.
4. Bezdushnyy, A.A., Bezdushnyy, A.N., Serebryakov, V.A., Filippov, V.I. (2005) *Metadata Integration of the Unified Scientific Information Space of RAS*. Moscow, 238 p. (in Russ.).
5. Kalenov, N.E., Pogorelko, K.P., Sotnikov, A.N. (2022) 'On the evolution of the digital library "Scientific heritage of Russia" as a component of the common digital space of scientific knowledge', *Information Processes*, 22(3) pp 155–166 (in Russ.).
6. Lin, J., Nogueira, R., Yates, A. (2022) *Pretrained Transformers for Text Ranking: Bert and Beyond*. Springer Cham Publ., 307 p. doi: 10.1007/978-3-031-02181-7.
7. Hjørland, B. (2022) 'Information retrieval and knowledge organization: a view from the point of view of the philosophy of science', *International Information Forum*, 47(4), pp. 36–57 (in Russ.). doi: 10.36535/0203-6460-2022-04-3.
8. Robinson, I., Webber, J., Eifrem, E. (2015) *Graph Databases*. O'Reilly Publ., CA, 220 p.
9. Pogorelko, K.P. (2020) 'A new version of the software for the electronic library "scientific heritage of Russia"', *Information Resources of Russia*, (5), pp. 27–29 (in Russ.).
10. Kalenov, N.E., Sotnikov, A.N. (2023) 'On the structure of the Common Digital Space of scientific knowledge ontology', *Sci. and Tech. Information. Ser. 2*, (7), pp. 20–26 (in Russ.). doi: 10.36535/0548-0027-2023-07-3.

Авторы**Погорелко Константин Павлович**^{1,2},

к.т.н., старший научный сотрудник,

konstpog@yandex.ru

Савин Геннадий Иванович^{1,2},

д.ф.-м.н., профессор, академик РАН,

руководитель научного направления,

savin@jscc.ru

Authors**Konstantin P. Pogorelko**^{1,2},

Cand. of Sci. (Engineering),

Senior Researcher, konstpog@yandex.ru

Gennady I. Savin^{1,2},

Dr.Sci. (Physics and Mathematics),

Professor, Academician of RAS,

Head of the Research Area, savin@jscc.ru

¹ Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия² Национальный исследовательский центр

«Курчатовский институт», г. Москва, 123182, Россия

¹ Joint Supercomputer Center of RAS,

Moscow, 119334, Russian Federation

² National Research Centre "Kurchatov Institute",

Moscow, 123182, Russian Federation