ЭЛЕКТРОННЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ

УДК 025.43:004.65ГПНТБ
https://doi.org/10.33186/1027-3689-2022-12-85-103

Применение тезаурусов при обработке поисковых запросов: от локального использования – к связанным данным

М. В. Гончаров[1], К. А. Колосов[2], Е. Ф. Бычкова[3]

[1, 2, 3]ГПНТБ России, Москва, Российская Федерация
[1, 2]Московский государственный лингвистический университет, Москва, Российская Федерация

[1]goncharov@gpntb.ru
[2]kolosov@gpntb.ru
[3]bef@gpntb.ru

Аннотация. Тематический поиск на естественном языке является самым сложным из-за фразеологической неоднозначности. Для решения этой проблемы информационные системы могут задействовать термины из контролируемых словарей, таких как тезаурусы. Классификации, тезаурусы, системы предметных рубрик, нормативные (авторитетные) файлы рассматриваются в открытом сетевом пространстве в среде связанных открытых данных (Linked Open Data, LOD) как LOD-словари. Связи между ними позволяют обогащать (дополнять) запросы пользователей словами из других словарей и программно переходить к поиску ресурсов в системах других библиотек. В статье рассматриваются возможности практического применения тезаурусов EUROVOC и GEMET с целью расширения поисковых запросов пользователей Единого открытого архива информации ГПНТБ России (ЕОАИ), Портала электронной библиотеки (ПЭБ) Парламентской библиотеки Федерального собрания Российской Федерации и тематической базы данных «Экология: наука и технологии», записи которых могут стать потенциально связанными. В настоящей работе приводятся результаты исследования и описан характер выявленных проблем.

Статья подготовлена в рамках Государственного задания «Информационное обеспечение научных исследований учёных и специалистов на базе Открытого архива ГПНТБ России как системы агрегации научных знаний (FNEG-2022-003)» на 2022–2024 гг.

Ключевые слова: тезаурус, семантический веб, связанные данные, Linked Open Data, LOD, экологическая информация, EUROVOC, GEMET, формат SKOS

Для цитирования: Гончаров М. В., Колосов К. А., Бычкова Е. Ф. Применение тезаурусов при обработке поисковых запросов: от локального использования – к связанным данным / М. В. Гончаров, К. А. Колосов, Е. Ф. Бычкова // Научные и технические библиотеки. 2022. № 12. С. 85–103. https://doi.org/10.33186/1027-3689-2022-12-85-103

DIGITAL INFORMATION RESOURCES

UDC 025.43:004.65ГПНТБ
https://doi.org/10.33186/1027-3689-2022-12-85-103

Applying thesauruses in expanding user search queries: From local use to linked data

Mikhail V. Goncharov[1], Kirill A. Kolosov[2] and Elena F. Bychkova[3]

[1, 2, 3]Russian National Public Library for Science and Technology, Moscow, Russian Federation
[1, 2]Moscow State Linguistic University, Moscow, Russian Federation

[1]goncharov@gpntb.ru
[2]kolosov@gpntb.ru
[3]bef@gpntb.ru

Abstract. The subject search in natural languages is the most difficult one due to phraseological ambiguities. To solve the problem, the information systems mobilize the terms in controlled dictionaries, e. g. thesauruses. The authors examine the classifications, thesauruses, subject headings, normative (authority) files within the context of the open networked space of the Linked Open Data environment (LOD). These links enable to enhance (complement) user queries with the words from other dictionaries, and to navigate through the other libraries' systems for the resources. The authors explore the possibility of practical application of EUROVOC and GEMET thesauruses to expand search queries initiated by the users of RNPLS&T's Single Open Information Archive (SOIA), Portal of Electronic Library (PEL) of the Parliamentary Library of the RF Federal Assembly and the thematic database "Ecology: Science

and technologies", which records could be potentially linked. The authors cite the study findings and characterize the problems revealed.

The article is prepared within the framework of the Government Order "Information support of scientific research of scientists and specialists on the basis of the RNPLS&T Open Archive as the scientific knowledge aggregation system, (FNEG-2022-003)" for the years 2022–2024.

Keywords: thesaurus, semantic web, linked data, Linked Open Data, LOD, ecological information, EUROVOC, GEMET, SKOS format

Cite: Goncharov M. V., Kolosov K. A., Bychkova E. F. Applying thesauruses in expanding user search queries: From local use to linked data / M. V. Goncharov, K. A. Kolosov, E. F. Bychkova // Scientific and technical libraries, 2022. No. 12. P. 85–103. https://doi.org/10.33186/1027-3689-2022-12-85-103

Введение

Среди перспективных технологий развития информационного поиска, в том числе для использования на веб-порталах библиотек, внимание исследователей привлекает технология семантического поиска в открытом сетевом пространстве (семантический веб), в котором классификационные схемы и тезаурусы могут рассматриваться как словари связанных данных [**1**]. Использование тезаурусов для расширения поисковых запросов изучается как российскими, так и зарубежными учёными [**2**, **3**]. В частности, результаты эксперимента по применению Общественно-политического тезауруса для поиска с расширением по дереву понятий были приведены в публикации [**4**]. При выполнении подавляющего числа запросов количество документов, найденных с использованием деревьев тезауруса, значительно превышало количество документов, найденных по словам. Полнота поиска с использованием деревьев тезауруса значительно возросла. Однако увеличение полноты поиска часто сопровождается снижением точности, то есть релевантным считается большее количество нерелевантных документов.

Более проработанным как с научной, так и с практической точек зрения является проект Российской государственной библиотеки (РГБ) по представлению Библиотечно-библиографической классификации

(ББК) в виде сетевой системы организации знаний, целью которого является достижение интероперабельности со словарями связанных данных (Linked Data vocabularies) в среде Глобальной сети. Как отмечают авторы публикации [**5**], в РГБ автоматический учёт смысловых отношений при поиске в электронном каталоге и электронной библиотеке обеспечен с самого начала их создания путём ввода непосредственно в библиографические записи словесных формулировок индексов. В конечном счёте РГБ реализовала проект Классификационной системы организации знаний на основе ББК и опубликовала эту систему в среде связанных открытых данных (Linked Open Data, LOD) (https://lod.rsl.ru). Поддержка семантического поиска в сетевом пространстве (Semantic Web) обеспечивается благодаря наличию словарей связанных данных. Классификации, тезаурусы, системы предметных рубрик, нормативные (авторитетные) файлы рассматриваются в открытом сетевом пространстве как LOD-словари. Связи между ними позволяют обогащать (дополнять) запросы пользователей словами из других словарей и программно переходить к поиску ресурсов в системах других библиотек. Так, в РГБ исследуются возможности установления связей между LOD-словарями ББК, УДК и MeSH [**5**].

LOD часто рассматривается как виртуальное облако, где каждый может получить доступ к любым данным, которые ему разрешено просматривать, а также добавлять любые данные, не нарушая исходный источник. Это обеспечивает открытую среду, в которой данные могут создаваться, подключаться и использоваться в масштабах интернета.

Однако процесс связывания наборов данных может быть очень сложным из-за различий форматов, структуры, семантики и понятийных меток. Самый известный проект в этом направлении – Инициатива по оценке выравнивания онтологий (Ontology Alignment Evaluation Initiative – OAEI), существующий с 2004 г. [**6**]. В 2022 г. в рамках данной инициативы рабочим группам предлагается к рассмотрению 11 проектов различной тематики, в том числе по медицине (anatomy), материаловедению и инженерии (mse), биоразнообразию и экологии (biodiv). Целью работы является поиск соответствия между различными онтологиями, используемыми в определённой области знаний. Так, в проекте biodiv сопоставляются онтология окружающей среды (Environment Ontology – ENVO) и онтология семантической сети для

технологий Земли и окружающей среды (Semantic Web for Earth and Environment Technology Ontology – SWEET). На странице рабочей группы по данному направлению (http://oaei.ontologymatching.org/2022/biodiv/index.html) отмечается, что в 2020 г. к исследованию присоединился проект AgroPortal (http://agroportal.lirmm.fr/), чтобы включить в работу дополнительные задачи сопоставления тезаурусов, в частности Общего многоязычного экологического тезауруса (GEneral Multilingual Environmental Thesaurus – GEMET) с тезаурусом «Анализ и экспериментирование с экосистемами» (Analysis and Experimentation on Ecosystems thesaurus – ANAEETHES). Другая цель связана с сопоставлением тезауруса в области сельского хозяйства AGROVOC (http://agroportal.lirmm.fr/ontologies/AGROVOC) и тезауруса Национальной сельскохозяйственной библиотеки США (US National Agricultural Library Thesaurus – NALT).

С точки зрения технологии связанных данных важным моментом является публикация тезаурусов в интернете в формате SKOS – Simple Knowledge Organization System («простая система организации знаний») [**7**]. Важной особенностью формата SKOS является поддержка языковых тегов для лексических терминов. То есть тезаурусы, представленные в формате SKOS, являются многоязычными, что даёт дополнительные возможности их использования для расширения поисковых запросов [**8**].

Установление связей между терминами тезаурусов имеет большое значение для технологии связанных данных, поскольку если, например, термины в тезаурусе AGROVOC связаны с терминами в тезаурусе GEMET, то все документы, проиндексированные в репозиториях документов, использующих AGROVOC или GEMET, также потенциально связаны [**9**].

Тезаурусы, являющиеся предметом исследования

Среди проектов информационных систем библиотек, как было отмечено в публикации [**10**], мы используем портал Единого открытого архива информации ГПНТБ России (ЕОАИ) [**11**] и Портал электронной библиотеки (ПЭБ) Парламентской библиотеки Федерального собрания Российской Федерации в качестве базовых площадок для исследования и апробации новых технологических решений. Для анализа воз-

можностей практического применения тезаурусов с целью расширения поисковых запросов пользователей в информационных системах ЕОАИ и ПЭБ более подробно рассмотрим два связанных тезауруса: EUROVOC и GEMET.

Тезаурус EUROVOC [**12**] разработан и управляется по инициативе и при поддержке Европейского Союза (ЕС). Это многоязычный многодисциплинарный тезаурус, охватывающий деятельность ЕС. Он содержит термины на 24 языках стран Европы. Тезаурус доступен в формате SKOS и поддерживается с использованием семантической веб-технологии. В 1998 г. в Парламентской библиотеке (ПБ) Федерального собрания Российской Федерации подготовили русскоязычную версию тезауруса EUROVOC, осуществили его перевод и адаптацию, подготовили издание многоязычного представления тезауруса. В 2000 г. было выполнено поэтапное объединение российской версии с собственным тезаурусом библиотеки. Объём полученного тезауруса оказался более 13 тыс. лексических единиц и около 24 тыс. семантических связей. Поиск с использованием терминов тезауруса возможен как по распознанным копиям, так и по атрибутам электронных документов, хранящимся в ПЭБ [**13**]. При этом, как отмечается в [**9**], тезаурус EUROVOC в формате SKOS (https://lod-cloud.net/dataset/eurovoc) содержит около 7 тыс. лексических единиц (концепций).

Общий многоязычный тезаурус окружающей среды (GEneral Multilingual Environmental Thesaurus) GEMET [**14**] охватывает тематику области и доступен на 29 языках. Данный тезаурус издаётся и управляется Европейской сетью экологической информации и наблюдений (European Environment Information and Observation Network). Тезаурус доступен в формате связанных данных, его SKOS-версия (https://lod-cloud.net/dataset/gemet) содержит более 5 тыс. лексических единиц (концепций) [**9**]. Анализ структуры файла SKOS-версии тезауруса GEMET показал, что его концепты включают связанные данные следующих типов:

термины тезауруса AGROVOC;

термины тезауруса EUROVOC;

термины из базы знаний DBpedia;

термины из Википедии;

термины экологического тезауруса UMTHES.

Тезаурус GEMET представляет интерес для исследования, поскольку в ГПНТБ России полномасштабно с 2003 г., а по факту с 1998 г. ведётся тематическая БД «Экология: наука и технологии» (БД «Экология») [**15**]. Источниками пополнения БД «Экология» являются российские журналы, поступающие в ГПНТБ России, – всего около 1 270 наименований. Из них выбирают статьи, которые могут представлять интерес для читателей научно-технической библиотеки. Около 30% библиографических записей составляют книги. В месяц создаётся около 300 аннотированных записей (включая монографии, отраслевые сборники, труды конференций и раритетные издания). По состоянию на ноябрь 2022 г. в этой БД насчитывается более 80 тыс. записей. Каждая библиографическая запись содержит ключевые слова, а также тематические рубрики. При формировании поискового словаря ключевые слова, помимо терминов из отдельного поля «ключевые слова», добавляются из заглавия, наименования серии и заглавия подсерии, сведений об ответственности, оригинала перевода, издания, в котором опубликована статья, и – что очень важно – аннотации. Таким образом формируется массив ненормированных ключевых слов, который частично пересекается с ключевыми словами, представленными в тезаурусах. То же самое можно сказать о ключевых словах, которые вводятся в специальное поле: это, как правило, авторские ключевые слова или же выделенные библиографами при создании библиографической записи (обязательным условием является несовпадение их со словами, автоматически генерируемыми из вышеперечисленных полей).

Для построения семантических связей с внешними ресурсами необходимо определить соответствие введённых ключевых слов тем, которые используются в связываемых ресурсах, например, за счёт использования нормированных ключевых слов, рубрик УДК, ГРНТИ, ББК и др. Поскольку в настоящее время в мире существует и используется некоторое количество международных тезаурусов, важно провести их анализ и сформулировать критерии, по которым можно выбрать один из них для формирования БД «Экология», а также учесть, при создании каких ресурсов и репозиториев эти тезаурусы применяются.

Задачи и цели исследования

Задачи исследования – разработать модель информационной системы для формирования соответствий между тематическими рубриками и ключевыми словами БД «Экология» и терминами тезауруса GEMET, установить соответствия (сопоставить данные), проанализировать связи между терминами тезаурусов GEMET и EUROVOC.

Цель исследования – показать, что тезаурус GEMET можно использовать для расширения поисковых запросов к БД «Экология», в том числе за счёт связей терминов данного тезауруса с тезаурусом EUROVOC. В результате построения предложенных соответствий записи БД «Экология» могут стать потенциально связанными с записями в БД ПЭБ.

Использование терминов GEMET в качестве ключевых слов

В публикации [**16**] описан подход, использованный для реализации наборов пространственных метаданных, при котором по крайней мере одно ключевое слово в описании ресурса должно быть предоставлено из Общего многоязычного тезауруса окружающей среды (GEMET). Следует отметить, что GEMET – это больше, чем набор ключевых слов и их переводов: он представляет собой таксономию понятий, связанных друг с другом более широким термином, более узким термином, и содержит родственные отношения между терминами. При выборе ключевого слова предлагается использовать несколько подходов:

Ведение статистики использования ключевых слов. Для каждого термина показывается счётчик его использования в качестве ключевого слова на этапе создания записи. При таком подходе формируется рейтинг использования ключевых слов, не связанный со структурой GEMET.

Использование структуры GEMET. На этапе заполнения ключевых слов создатель записи получает древовидное представление терминов GEMET. Появляется возможность выбора одного или нескольких дополнительных ключевых слов среди более узких терминов. Этот подход не зависит от того, как часто ключевое слово использовалось ранее.

Комбинированный подход, основанный на сочетании структурированного и статистических подходов.

Как было отмечено выше, при наличии связей между рубриками и ключевыми словами БД «Экология» и концепциями GEMET записи БД «Экология» становятся потенциально связанными с записями в БД Парламентской электронной библиотеки (БД ПЭБ). Кроме того, для расширения поискового запроса можно использовать как лексические единицы GEMET, так и все привязанные к ним термины других источников, например, Википедии или БД DBpedia. На рис. 1 приведена общая схема идеи построения связей между библиографическими записями БД «Экология» и другими информационными источниками.

Модель информационной системы

Для реализации предложенной модели построения связей с использованием тезауруса GEMET была разработана БД, содержащая концепты данного тезауруса, с возможностью создания реляционных связей с терминами и ключевыми словами БД «Экология». На рис. 2 представлена функциональная схема расширенной обработки поискового запроса к БД «Экология» с использованием тезауруса GEMET.

Проведённый программный анализ записей БД «Экология» показал, что в ней используется около 4 800 рубрик, среди которых потенциально связанными с терминами тезауруса GEMET являются 2 240 рубрик.

Среди используемых в БД «Экология» приблизительно 6 500 ключевых слов потенциально связанными с терминами тезауруса GEMET являются 454 ключевых слова. Потенциально связанными считались рубрики и ключевые слова, формулировки которых были найдены в определениях концептов тезауруса GEMET.

Анализ наличия рубрик и ключевых слов в библиографических записях БД «Экология» показал, что рубрики проставлены приблизительно в 20% записей, ключевые слова – в 5% записей, при этом число записей, в которых присутствуют одновременно рубрики и ключевые слова, составляет менее 0,5%. Большинство библиографических записей БД «Экология» содержит несколько рубрик или ключевых слов, что повышает вероятность их использования при построении связей на основе терминов GEMET.

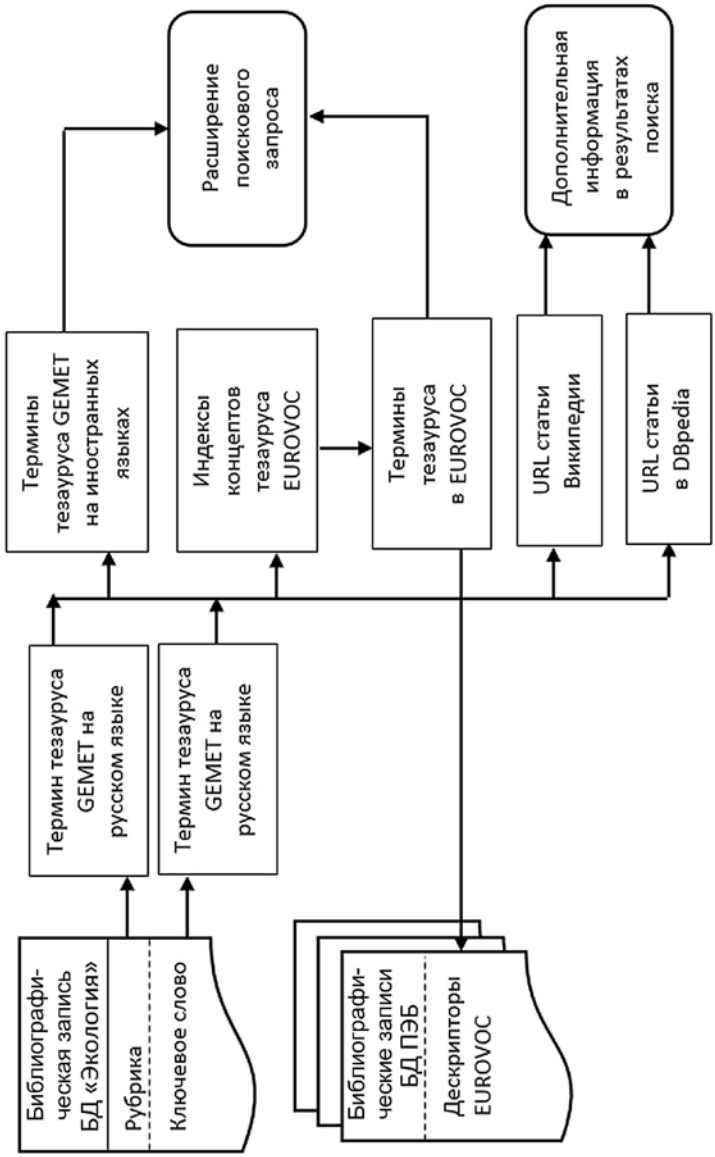

Рис. 1. Принцип построения связей между библиографическими записями БД «Экология» и другими информационными источниками

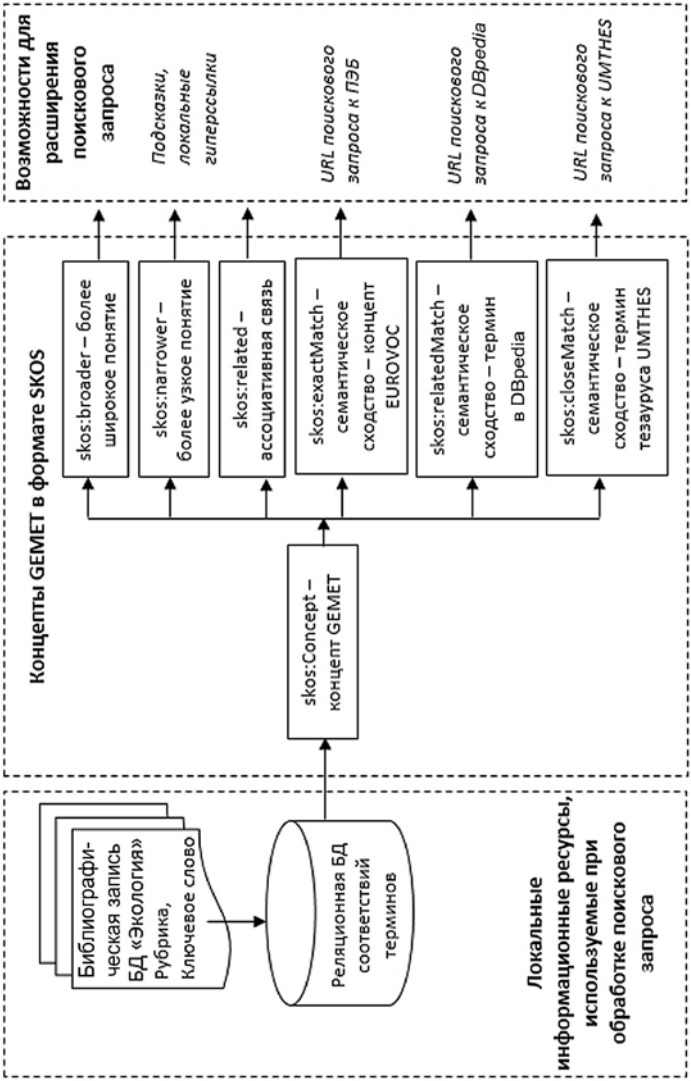

Рис. 2. Функциональная схема расширенной обработки поискового запроса к БД «Экология» с использованием тезауруса GEMET

Принцип построения реляционных связей в рассматриваемой модели

Принцип построения реляционных связей ключевых слов БД «Экология» с концептами тезауруса GEMET покажем на примере термина «биотехнология». По данному ключевому слову в БД «Экология» найдено 220 библиографических записей.

Поиск по слову «биотехнология» в БД тезауруса GEMET выдаёт следующие результаты:

биотехнология (concept/898) → Группа → ИССЛЕДОВАНИЯ, НАУКА → Супергруппа → СОЦИАЛЬНЫЕ АСПЕКТЫ, МЕРЫ ЭКОЛОГИЧЕСКОЙ ПОЛИТИКИ

Соединение биологии и технологии. Термин используется для описания достижений в области использования биологических организмов в научных и коммерческих целях. «Био» означает биология и наука о жизни, а «технология» – приёмы и инструменты, которые биотехнологи имеют на своём вооружении. К этим инструментам относятся микроорганизмы и широкий спектр приёмов манипулирования с ними, к которым относится, в частности, генная инженерия.

биотехнологическая угроза (concept/12814) → Группа → РИСКИ, БЕЗОПАСНОСТЬ → Супергруппа → СОЦИАЛЬНЫЕ АСПЕКТЫ, МЕРЫ ЭКОЛОГИЧЕСКОЙ ПОЛИТИКИ

сельскохозяйственная биотехнология (concept/185) → Группа → ИССЛЕДОВАНИЯ, НАУКА → Супергруппа → СОЦИАЛЬНЫЕ АСПЕКТЫ, МЕРЫ ЭКОЛОГИЧЕСКОЙ ПОЛИТИКИ

биотехнология в области здоровья (concept/3880) → Группа → ИССЛЕДОВАНИЯ, НАУКА → Супергруппа → СОЦИАЛЬНЫЕ АСПЕКТЫ, МЕРЫ ЭКОЛОГИЧЕСКОЙ ПОЛИТИКИ

новые продукты питания (concept/5701) → Группа → ПРОДУКТЫ, МАТЕРИАЛЫ → Супергруппа → ДЕЯТЕЛЬНОСТЬ ЧЕЛОВЕКА И ЕЁ РЕЗУЛЬТАТЫ, ВОЗДЕЙСТВИЕ НА ОКРУЖАЮЩУЮ СРЕДУ

социально-экономический эффект от применения биотехнологий (concept/7827) → Группа → ВОЗДЕЙСТВИЕ, ВЛИЯНИЕ → Супергруппа → ДЕЯТЕЛЬНОСТЬ ЧЕЛОВЕКА И ЕЁ РЕЗУЛЬТАТЫ, ВОЗДЕЙСТВИЕ НА ОКРУЖАЮЩУЮ СРЕДУ

Из найденных концептов термину «биотехнология» соответствует concept/898 ***(биотехнология)***. Среди близких к данному термину в БД «Экология» имеются рубрики: *«Биотехнология – Окружающая среда»* (шесть записей в БД «Экология») и *«Биотехнология – Применение в сельском хозяйстве»* (одна запись в БД «Экология»), которым соответствуют concept/7827 и concept/185 тезауруса GEMET. В табл. 1 показан пример расширения поискового запроса за счёт связи между терминами тезаурусов GEMET и EUROVOC.

Поскольку тезаурус EUROVOC используется в странах Европы, то можно искать документы с использованием его терминов на портале общедоступных документов ЕС – EUR-Lex (https://eur-lex.europa.eu). Важно заметить, что информация, отображаемая в EUR-Lex, извлекается из общего хранилища метаданных и контента The semantic repository of the Publications Office – Cellar. Cellar включает около 150 млн документов на 24 языках. Метаданные на таком же количестве языков хранятся и описываются в RDF, в результате чего получается около 800 млн триплетов [**17**]. Cellar базируется на технологиях семантического веба, в рамках нескольких стандартов, таких как RDF Schema и OWL. Данные в Cellar не ограничиваются юридическими документами и информацией, опубликованными в EUR-Lex, но также включают другие публикации ЕС (книги, брошюры и др.). К данным, хранящимся в Cellar, можно получить прямой доступ через конечную точку SPARQL и RESTful API. EUR-Lex предлагает веб-сервис, который позволяет искать и извлекать метаданные из репозитория. Любой поиск, сделанный на веб-сайте, также может быть выполнен через веб-службу, включая полнотекстовый поиск.

В табл. 2 показана связь найденных концептов тезауруса GEMET с терминами и статьями в DBpedia, Википедии, а также терминами тезауруса UMTHES на примере поиска по термину «биотехнология».

Таблица 1

Расширение результатов поиска по термину «биотехнология» с использованием связи между тезаурусами GEMET и EUROVOC

Число найденных записей в БД «Экология»	Концепт тезауруса GEMET	Связанный термин тезауруса EUROVOC	Число найденных записей в БД ПЭБ по связанному термину тезауруса EUROVOC	Число найденных записей на портале EUR-Lex (https://eur-lex.europa.eu) по связанному термину тезауруса EUROVOC
220	*биотехнология* (concept/ 898)	*http://eurovoc. europa.eu/3797* Сведения из тезауруса ПБ: *Биотехнология* – дескриптор Уровень: 3 Вышестоящий: ТЕХНОЛОГИИ Термин английский: *BIOTECHNOLOGY* Термин французский: *BIOTECHNOLOGIE*	517	2 539

Таблица 2

Расширение результатов поиска по термину «биотехнология» с использованием связи между тезаурусами GEMET и другими информационными источниками

Концепт тезауруса GEMET	Связанный термин экологического тезауруса UMTHES	URL для доступа к информации из БД DBpedia	URL для доступа к информации из Википедии
Биотехнология (concept/898)	https://sns.uba.de/umthes/de/ concepts/_00005093.html	http://dbpedia.org/resource/Biotechnology	https://en.wikipedia.org/wiki/Biotechnology
Социально-экономический эффект от применения биотехнологий (concept/7827)	https://sns.uba.de/umthes/de/concepts/_00051229.html	Связь не определена	Связь не определена
Сельскохозяйственная биотехнология (concept/185)	https://sns.uba.de/umthes/de/concepts/_00051528. html	http://dbpedia.org/resource/ Agricultural_science	http://en.wikipedia.org/wiki/Agricultural_science

Заключение

На основе изложенного принципа построения связей между библиографическими записями БД «Экология» и другими информационными источниками в настоящее время продолжается работа по построению реляционных связей ключевых слов и рубрик БД «Экология» с концептами тезауруса GEMET. После окончания этого этапа на модели ЕОАИ будет размещён интерфейс с возможностями расширения поисковых запросов пользователей на основе применения связанных данных, базирующийся на работе с тезаурусами GEMET и EUROVOC. Следует заметить, что, несмотря на относительно небольшой объём, аналитическая БД «Экология» является востребованным ресурсом у читателей ГПНТБ России. Коме того, её содержимое вливается в основной электронный каталог, доступный через веб-портал библиотеки (https://cat.gpntb.ru). Использование технологии связанных данных позволяет обогатить результаты поиска информации, привлечь новые информационные источники.

Одной из проблем использования ключевых слов для расширения поисковых запросов является отсутствие практики использования контролируемого словаря ключевых слов на этапе создания библиографических записей. Это ограничивает возможности для связи ключевых слов с терминами тезауруса и, как следствие, не позволяет найти ресурсы при поиске с использованием синонимов, более широких или более узких терминов, а также обрабатывать поисковые запросы на иностранных языках.

В данной статье был рассмотрен вариант использования связей между терминами предметно ориентированного тезауруса с тематической БД, когда пересечение ключевых слов и предметных рубрик с концептами тезауруса можно установить достаточно быстро. Дальнейшее исследование будет связано с анализом результативности внедрения предложенной модели. Задача по установлению семантического соответствия между универсальными классификационными схемами, как было отмечено в [**5**], решается далеко не так просто. Тем не менее современные тенденции науки – это развитие технологии семантического поиска, создание систем организации знаний и расширение представлений классификационных схем в Семантической паутине.

Список источников

1. **Малахов Д. А., Серебряков В. А.** Модель семантического поиска на базе тезауруса // CEUR Workshop Proceedings. 2017. Vol. 2022. P. 191–196.

2. **Атаева О. М., Серебряков В. А., Тучкова Н. П.** Расширение предметной области информационного запроса на основе онтологии знаний цифровой библиотеки LibMeta // Научный сервис в сети Интернет. Федеральное государственное учреждение «Федеральный исследовательский центр Институт прикладной математики им. М. В. Келдыша Российской академии наук», 2019. Т. 21. С. 63–75.

3. **Kechagioglou X. et al.** EcoPortal: An Environment for FAIR Semantic Resources in the Ecological Domain // Proceedings. 2021. Т. 1613. С. 0073. URL: http://ceur-ws.org.

4. **Добров Б. В., Лукашевич Н. В**. Тезаурус РуТез как ресурс для решения задач информационного поиска // Знания – Онтологии – Теории – 2009. URL: http://ns.math.nsc.ru/conference/zont09/reports/93Dobrov-Lukashevich.pdf (дата обращения: 01.12.2022).

5. **Лаврёнова О. А., Винберг А. А.** Современные пользователи библиотек и пространство связанных открытых данных // Библиотековедение. 2020. Т. 69. № 3. С. 243–260.

6. **Ontology** Alignment Evaluation Initiative. URL: http://oaei.ontologymatching.org (дата обращения: 12.11.2022).

7. **SKOS** Simple Knowledge Organization System Reference.
URL: https://www.w3.org/TR/skos-reference (дата обращения: 12.11.2022).

8. **Rodríguez-Enríquez C. A. et al.** Supply chain knowledge management: A linked data-based approach using SKOS // Dyna. 2015. Vol. 82. № 194. P. 27–35.

9. **Morshed A., Caracciolo C., Johannsen G., Keizer J.** Thesaurus alignment for Linked Data publishing. In: Proceedings of the International Conference on Dublin Core and Metadata Applications 2011. P. 37–46. Dublin Core Metadata Initiative.

10. **Гончаров М. В., Колосов К. А.** Проблемы релевантности при обработке поисковых запросов к библиографическим и полнотекстовым базам данных в современных моделях обеспечения научных исследований средствами открытых архивов // Научные и технические библиотеки. 2022. № 11. С. 120–134.

11. **Шрайберг Я. Л., Гончаров М. В., Колосов К. А**. О разработке концепции Открытого архива информации ГПНТБ России // Научные и технические библиотеки. 2020. № 12. С. 45–58.

12. **EuroVoc**: Википедия. Свободная энциклопедия.
URL: https://en.wikipedia.org/wiki/EuroVoc (дата обращения: 12.11.2022).

13. **Москаленко Т. А., Мякова Н. А**. Информационно-поисковый тезаурус Парламентской библиотеки: этапы разработки, ведение, применение и дальнейшие перспективы // Научные и технические библиотеки. 2009. № 3. С. 18–22.

14. **GEMET** – GEneral Multilingual Environmental Thesaurus.
URL: https://www.eionet.europa.eu/gemet/en/about (дата обращения: 12.11.2022).

15. **Боргоякова К. С., Бычкова Е. Ф., Земсков А. И., Кондрашева И. Ю.** Библиометрический анализ научных публикаций по экологии на основе реферативной базы данных «Экология: наука и технологии» ГПНТБ России // Научные и технические библиотеки. 2017. № 10. С. 54–68.

16. **Ostländer N., Lutz M.** INSPIRE-ing GEMET-Enhancing Metadata Creation and Discovery // EnviroInfo. 2008. P. 212–214.
URL: http://enviroinfo.eu/sites/default/files/pdfs/vol119/0212.pdf
(дата обращения: 05.12.2022).

17. **Francesconi E.** On the future of legal publishing services in the Semantic Web // Future Internet. 2018. Vol. 10. № 6. P. 48.

References

1. **Malahov D. A., Serebriakov V. A.** Model` semanticheskogo poiska na baze tezaurusa // CEUR Workshop Proceedings. 2017. Vol. 2022. P. 191–196.

2. **Ataeva O. M., Serebriakov V. A., Tuchkova N. P.** Rasshirenie predmetnoi` oblasti informatcionnogo zaprosa na osnove ontologii znanii` tcifrovoi` biblioteki LibMeta // Nauch-

ny`i` servis v seti Internet. Federal`noe gosudarstvennoe uchrezhdenie «Federal`ny`i` issledovatel`skii` centr Institut pricladnoi` matematiki im. M. V. Keldy`sha Rossii`skoi` akademii nauk», 2019. T. 21. S. 63–75.

3. **Kechagioglou X. et al.** EcoPortal: An Environment for FAIR Semantic Resources in the Ecological Domain // Proceedings. 2021. T. 1613. C. 0073. URL: http://ceur-ws.org.

4. **Dobrov B. V., Lukashevich N. V.** Tezaurus RuTez kak resurs dlia resheniia zadach informatcionnogo poiska // Znaniia – Ontologii – Teorii – 2009. URL: http://ns.math.nsc.ru/conference/zont09/reports/93Dobrov-Lukashevich.pdf (data obrashcheniia: 01.12.2022).

5. **Lavryonova O. A., Vinberg A. A.** Sovremenny`e pol`zovateli bibliotek i prostranstvo sviazanny`kh otkry`ty`kh danny`kh // Bibliotekovedenie. 2020. T. 69. № 3. S. 243–260.

6. **Ontology** Alignment Evaluation Initiative. URL: http://oaei.ontologymatching.org (data obrashcheniia: 12.11.2022).

7. **SKOS** Simple Knowledge Organization System Reference. URL: https://www.w3.org/TR/skos-reference (data obrashcheniia: 12.11.2022).

8. **Rodríguez-Enríquez C. A. et al.** Supply chain knowledge management: A linked data-based approach using SKOS // Dyna. 2015. Vol. 82. № 194. P. 27–35.

9. **Morshed A., Caracciolo C., Johannsen G., Keizer J.** Thesaurus alignment for Linked Data publishing. In: Proceedings of the International Conference on Dublin Core and Metadata Applications 2011. P. 37–46. Dublin Core Metadata Initiative.

10. **Goncharov M. V., Kolosov K. A.** Problemy` relevantnosti pri obrabotke poiskovy`kh zaprosov k bibliograficheskim i polnotekstovy`m bazam danny`kh v sovremenny`kh modeliakh obespecheniia nauchny`kh issledovanii` sredstvami otkry`ty`kh arhivov // Nauchny`e i tekhnicheskie biblioteki. 2022. № 11. S. 120–134.

11. **Shrai`berg Ia. L., Goncharov M. V., Kolosov K. A**. O razrabotke kontceptcii Otkry`togo arhiva informatcii GPNTB Rossii // Nauchny`e i tekhnicheskie biblioteki. 2020. № 12. S. 45–58.

12. **EuroVoc**: Vikipediia. Svobodnaia e`ntciclopediia. URL: https://en.wikipedia.org/wiki/EuroVoc (data obrashcheniia: 12.11.2022).

13. **Moskalenko T. A., Miakova N. A**. Informatcionno-poiskovy`i` tezaurus Parlamentskoi` biblioteki: e`tapy` razrabotki, vedenie, primenenie i dal`nei`shie perspektivy` // Nauchny`e i tekhnicheskie biblioteki. 2009. № 3. S. 18–22.

14. **GEMET** – GEneral Multilingual Environmental Thesaurus. URL: https://www.eionet.europa.eu/gemet/en/about (data obrashcheniia: 12.11.2022).

15. **Borgoiakova K. S., By`chkova E. F., Zemskov A. I., Kondrasheva I. Iu.** Bibliometricheskii` analiz nauchny`kh publikatcii` po e`kologii na osnove referativnoi` bazy` danny`kh «E`kologiia: nauka i tekhnologii» GPNTB Rossii // Nauchny`e i tekhnicheskie biblioteki. 2017. № 10. S. 54–68.

16. **Ostländer N., Lutz M.** INSPIRE-ing GEMET-Enhancing Metadata Creation and Discovery // EnviroInfo. 2008. P. 212–214.
URL: http://enviroinfo.eu/sites/default/files/pdfs/vol119/0212.pdf (data obrashcheniia: 05.12.2022).

17. **Francesconi E.** On the future of legal publishing services in the Semantic Web // Future Internet. 2018. Vol. 10. № 6. P. 48.

Информация об авторах / Information about the authors

Гончаров Михаил Владимирович – канд. техн. наук, доцент, ведущий научный сотрудник, руководитель группы перспективных исследований и аналитического прогнозирования ГПНТБ России, доцент Московского государственного лингвистического университета, Москва, Российская Федерация

goncharov@gpntb.ru

Колосов Кирилл Анатольевич – канд. техн. наук, ведущий научный сотрудник ГПНТБ России, доцент Московского государственного лингвистического университета, Москва, Российская Федерация

kolosov@gpntb.ru

Бычкова Елена Феликсовна – ведущий научный сотрудник, руководитель группы развития проектов в области экологии и устойчивого развития ГПНТБ России, Москва, Российская Федерация

bef@gpntb.ru

Mikhail V. Goncharov – Cand. Sc. (Engineering), Associate Professor, Leading Researcher, Head, Group for Perspective Research and Analytic Forecasting, Russian National Public Library for Science and Technology; Associate Professor, Moscow State Linguistic University, Moscow, Russian Federation

goncharov@gpntb.ru

Kirill A. Kolosov – Cand. Sc. (Engineering), Leading Researcher, Russian National Public Library for Science and Technology; Associate Professor, Moscow State Linguistic University, Moscow, Russian Federation

kolosov@gpntb.ru

Elena F. Bychkova – Cand. Sc. (Pedagogy), Leading Researcher, Head, Ecology and Sustainable Development Group of Academic Secretary Department, Russian National Public Library for Science and Technology, Moscow, Russian Federation

bef@gpntb.ru