

ЕДИНОЕ ЦИФРОВОЕ ПРОСТРАНСТВО НАУЧНЫХ ЗНАНИЙ КАК ИНТЕГРАТОР ПОЛИТЕМАТИЧЕСКИХ ИНФОРМАЦИОННЫХ РЕСУРСОВ

© 2024 Н. Е. Каленов^{1,*}, А. Н. Сотников^{1,**}

Представлено академиком РАН Г.И. Савиным

Поступило 27.07.2023 г.

После доработки 22.01.2024 г.

Принято к публикации 29.01.2024 г.

Рассматриваются цели, задачи и структура онтологии Единого цифрового пространства научных знаний (ЕЦПНЗ). ЕЦПНЗ представляет собой интегрированную информационную структуру, объединяющую представленные в сети Интернет государственные научные информационные системы (Большую российскую энциклопедию, Национальную электронную библиотеку, Государственный каталог географических названий и др.) с отраслевыми информационными системами, базами данных и электронными библиотеками (MathNet, “Соционет”, “Научное наследие России” и др.). ЕЦПНЗ может рассматриваться в качестве информационной основы для решения задач искусственного интеллекта. В статье приводится разработанная в МСЦ РАН унифицированная структура онтологии ЕЦПНЗ и ее моделирование на примере 10 предметных и восьми вспомогательных классов объектов универсального подпространства ЕЦПНЗ.

Ключевые слова: информационные ресурсы, онтологический подход, пространство научных знаний, структуризация, контент, semantic Web

DOI: 10.31857/S2686954324010177, **EDN:** ZSHUOC

1. ЦЕЛИ И ЗАДАЧИ ЕЦПНЗ

Концепция Единого цифрового пространства научных знаний (ЕЦПНЗ) как структурированной информационной среды, отражающей научные достижения, была предложена в 2019 г. [1]. До этого высказывалось много соображений общего характера, касающихся необходимости создания “единого информационного пространства” [2, 3] или “электронного пространства знаний” [4]. Обсуждению целей, функций и архитектуры построения ЕЦПНЗ была посвящена Всероссийская научная конференция “Единое цифровое пространство знаний; проблемы и решения” [5], в которой приняла участие несколько десятков специалистов из научно-исследовательских организаций, вузов, научных библиотек и музеев.

Одной из главных целей создания ЕЦПНЗ является информационная поддержка научных исследований. Если раньше, в “эпоху печатных изданий”, существовали специальные службы (отраслевые центры и отделы научной информации в НИИ и академических библиотеках), обеспечивающие релевантной информацией ученых и специалистов, то сейчас развитие сетевых технологий и замена печатных изданий цифровыми породили тезис “в интернете всё есть”, следствием чего явилось перекладывание задачи информационного сопровождения научных исследований на плечи самих ученых. Этот подход приводит к тому, что исследователь тратит львиную долю своего времени и интеллектуальных усилий на ознакомление с информацией, значительная часть которой не соответствует его научным интересам, вместо того чтобы заниматься исследованиями. С другой стороны, если он этого не сделает, отсутствие информации может привести к дублированию исследований и бессмысленной затрате времени на получение результатов, ранее уже полученных другими исследователями.

¹Межведомственный суперкомпьютерный центр (МСЦ) РАН — филиал ФГУ ФНЦ Научно-исследовательский институт системных исследований (НИИСИ) Российской академии наук, Москва, Россия

*E-mail: nekalenov@mail.ru

**E-mail: asotnikov@jssc.ru

Создание ЕЦПНЗ, содержащего достоверную и не дублированную научную фактографическую и документальную структурированную информацию по различным разделам науки снимет в значительной мере остроту проблемы информационного сопровождения научных исследований. Но это лишь одна из целей создания ЕЦПНЗ.

Общая цель создания ЕЦПНЗ — формирование и поддержка цифровой информационной среды, необходимой для решения комплекса задач развития общества, обеспечивающих:

- информационное сопровождение научных исследований;
- поддержку образовательных процессов, начиная от средней школы (тесно связанных с популяризацией науки) и кончая аспирантурой (напрямую связанных с информационным сопровождением науки);
- популяризацию науки (развитие мотивации для занятий наукой и получения соответствующего образования, одобрение финансирования науки со стороны общества);
- сохранность научных знаний;
- процессы мониторинга и управления наукой.

Таким образом, ЕЦПНЗ должно выполнять научно-информационные, образовательные, общекультурные и управленческие функции.

Соответственно, ЕЦПНЗ должно содержать элементы, ориентированные на следующие категории пользователей:

- исследователи, которым должна быть предоставлена многоаспектная ретроспективная и текущая (отфильтрованная по различным критериям, обеспечивающим ее достоверность и новизну¹) информация по соответствующему их интересам научному направлению;
- школьники и студенты, которые должны получать проверенную временем достоверную базовую информацию различного уровня; эта информация должна включать факты со ссылками на учебные материалы, собственно тексты классических учебников (отбор их также должен осуществляться на основе независимых крите-

риев), цифровые модели тех или иных явлений и событий;

- специалисты-аналитики и представители управленческих структур, анализирующие состояние и тенденции развития различных областей науки;
- широкая общественность, которую необходимо знакомить с наиболее интересными результатами, полученными в той или иной области науки, а также с историей научных открытий и их авторами;
- специалисты и “любители”, интересующиеся историей науки и ее творцами.

Отличительной особенностью ЕЦПНЗ является то, что его программная оболочка, используя возможности искусственного интеллекта, в т.ч. по анализу естественного языка, должна обрабатывать широкий спектр политематических запросов, не обязательно содержащих термины, в явном виде присутствующие в метаданных, относящихся к конкретным объектам, отраженным в ЕЦПНЗ. Например, система должна отвечать на запросы типа: “какие археологические находки были обнаружены на Урале в 19 веке”. На этот запрос должны быть выданы описания всех археологических объектов, найденных в разных районах, относящихся к Уралу, за период с 1801 по 1900 гг. При этом в информации об отдельном объекте может содержаться указание лишь на конкретное место его обнаружения, а заключение о том, что данное место относится к Уралу, должно вытекать из автоматического анализа связей археологических объектов с объектами других категорий (в данном случае относящихся к географии и времени).

Результатом обращения к ЕЦПНЗ должен быть не только однозначный достоверный ответ на фактографический вопрос, но и предоставление возможности ознакомиться, не выходя за пределы “пространства”, с источником этого ответа и связанными с ним разнородными научными ресурсами.

В этой связи ЕЦПНЗ рассматривается как интегратор для научных целей государственных информационных систем (таких как Большая российская энциклопедия — БРЭ [6], Национальная электронная библиотека — НЭБ [7], Российский индекс научного цитирования — РИНЦ [8], Государственный каталог географических названий ГКГН [9]) с отраслевыми

¹ Фильтрация может осуществляться на основе критериев отбора источников информации, по экспертным оценкам признанных авторитетных ученых и т.п.

научными информационными системами, электронными библиотеками, регистрами и т.п.

Наиболее активные исследования, связанные с практическим построением ЕЦПНЗ, проводятся в межведомственном суперкомпьютерном центре (МСС) РАН — филиале ФГУ ФНЦ НИИСИ РАН. Здесь разработана архитектура ЕЦПНЗ, предложена структура его онтологии, ведутся исследования, связанные с реализацией конкретных решений, отрабатываемых, в частности, на примере электронной библиотеки “Научное наследие России” и ее информационных фондов [10].

2. СТРУКТУРА ЕЦПНЗ

Основными составляющими ЕЦПНЗ являются онтология и контент. Согласно предложенной модели онтологии [11], ЕЦПНЗ включает совокупность подпространств (ПП) по различным областям науки, построенных по единым принципам, основанным на использовании онтологических стандартов, принятых в среде Semantic WEB [12–15].

ЕЦПНЗ представляется в виде пятиуровневой иерархической структуры:

ЕЦПНЗ — подпространства — классы объектов — атрибуты объектов класса — значения атрибутов.

Основой контента ЕЦПНЗ являются объекты, которые представляют собой совокупность структурированных многоаспектных данных, отражающих информацию о физической сущности (например, о конкретном человеке, конкретной книге, музейном предмете и т.п.), научном понятии (например, о теореме Пифагора, уравнениях Максвелла, корпусе текстов японского языка и т.п.), событии и т.д.

Каждый объект ЕЦПНЗ характеризуется набором атрибутов (свойств) и их значений, а также связями с другими объектами. Перечень атрибутов объекта устанавливается исходя из роли, которую объект играет при решении задач в рамках ЕЦПНЗ.

Класс — совокупность объектов, имеющих заданный набор атрибутов.

Подпространство — совокупность классов объектов. ЕЦПНЗ включает универсальное и тематические подпространства.

Универсальное подпространство содержит вспомогательные классы объектов и предметные классы объектов мультидисциплинарного характера. Вспомогательные классы используются при описании научных объектов всех подпространств и определяют правила формирования атрибутов объектов предметных классов и связей между ними; к ним относятся единицы измерения, универсальные классификационные системы, форматы данных и т.п. К предметным классам универсального подпространства могут быть отнесены общенаучные события, ученые мирового класса, публикации и т.п.

Тематическое подпространство (например, подпространство “математика”, “информатика”, “космические исследования” и т.п.) содержит элементы, напрямую связанные с данным научным направлением, а также связи с элементами универсального и других тематических подпространств.

Единство пространства обеспечивается связями между парами объектов и значений их атрибутов.

Связи ЕЦПНЗ подразделяются на три типа — универсальные, квазиуниверсальные и специфические.

Связи могут быть простыми и составными. Простые связи содержат (в терминах триплетов RDF [16]) указание на субъект, объект и (факкультативно, в зависимости от конкретного вида связи) значение связи. Значения составных связей могут содержать “вложения” — иметь собственные атрибуты и их значения; количество вложений не ограничено и определяется справочником данной связи.

Универсальные связи являются простыми и указывают лишь на факт отношений между элементами и не зависят от классов объектов, которые они связывают. К этому классу относятся связи “эквивалентно”; “пересекается с”; “содержит”; “содержится в”.

Связи этого вида широко употребляются в предметных тезаурусах и при установлении соответствия между элементами классификационных систем. В ЕЦПНЗ они дополнительно используются, например, при указании соподчиненности подразделений организаций, различных наименований организаций, различного написания фамилий и имен персон и т.п.

Квазиуниверсальные связи связывают субъекты различных классов с объектами заданного

класса, они могут быть простыми или составными. Перечень квазиуниверсальных связей может пополняться по мере развития ЕЦПНЗ и добавления новых элементов. Примером квазиуниверсальных связей могут служить ссылки на статьи в энциклопедии, относящиеся к персоне, организации, событию, научному открытию и т.п.; временные характеристики, указывающие на начало или конец того или иного процесса или события.

Специфические связи устанавливаются между субъектами и объектами заданных классов; они могут быть простыми и составными. Количество и вид специфических связей определяются при формировании онтологий конкретных классов. В отличие от универсальных связей, которые имеют статичный характер, и квазиуниверсальных связей, набор которых растет достаточно медленно, перечень специфических связей является достаточно динамичным, поскольку определяется развитием ЕЦПНЗ и возникающими перед ним задачами. Специфические связи имеют атрибуты, значения которых, в свою очередь, могут иметь свои атрибуты со своими значениями. Например, связь “вид взаимоотношений персоны с организацией” имеет перечень атрибутов, один из которых – “сотрудник”; значением этого атрибута является наименование должности, которое может иметь факультативно один или два атрибута – дата начала и окончания работы в данной должности. К специфическим относится связь персоны с публикацией, которая может принимать значения “автор”, “переводчик”, “художник”, “составитель” и т.д. В качестве специфической, исходя из задач ЕЦПНЗ, может выступать связь между научным открытием и публикацией. Эта связь может рекомендовать ту или иную публикацию для различных групп пользователей; исходя из задач ЕЦПНЗ, она может принимать значения: “впервые опубликовано в”, “рекомендуемая научная монография”, “рекомендуемое учебное издание для вузов”, “рекомендуемый учебник для школ”, “рекомендуемое научно-популярное издание”².

Информация о структуре всех элементов ЕЦПНЗ отражается в справочниках, имеющих структуру, унифицированную для каждого из

шести видов элементов (подпространство, класс объектов, связи трех типов, атрибуты объектов/связей).

Конкретные значения атрибутов и связей каждого объекта содержатся в словарях, информация о которых отражается в соответствующих справочниках.

Справочники каждого уровня иерархии могут иметь различное количество составляющих, определяющих конкретные элементы данного вида. Исключение составляет базовый справочник верхнего уровня с именем CDSSK, который включает девять составляющих. Первые шесть описывают структуру справочников элементов, последние три – структуру словарей атрибутов, объектов и связей.

3. ФОРМАЛИЗОВАННОЕ ОПИСАНИЕ ОНТОЛОГИИ ЕЦПНЗ

Все элементы ЕЦПНЗ, включая справочники и словари, имеют свои уникальные структурированные имена (URN); предложена мнемоника формирования URN, позволяющая достаточно просто добавлять в ЕЦПНЗ новые атрибуты объектов и связи между ними.

Составляющие справочника ЕЦПНЗ имеют вид:

CDSSK.1: Структура справочника подпространств.

Имя (URN) справочника подпространств: SUBS.

Составляющие:

название подпространства;

префикс (два символа);

описание (пояснительный текст).

Примеры:

SUBS.1: Универсальное; UN; подпространство, включающее классы объектов, не связанные непосредственно с конкретной научной тематикой, в т.ч. универсальные справочные данные.

SUBS.2: Информатика; 20; подпространство включает объекты, относящиеся к научному направлению “информатика”.

CDSSK.2: Структура справочника классов.

² Очевидно, что рекомендации должны формироваться на основе экспертных оценок публикации, формируемых научным сообществом.

Имя (URN) справочника классов: Class.

Элементы:

название;

префикс (UNab для универсального и <ПП>ab для локального, где <ПП> префикс тематического подпространства — два символа; ab — два произвольных буквенно-цифровых символа);

URN словаря атрибутов;

описание (пояснительный текст).

Примеры:

Class 1: персоны; UN; UNPS; A_UNPS; информация о персонах, в той или иной мере связанных с научными исследованиями.

Class 2. политематические базы данных; UN; UNBD; A_UNBD; энциклопедии, БД персональных, организаций, документальные БД, каталоги ресурсов, электронные библиотеки.

Class 3: Форматы; UN; UNFT; A_UNFT; форматы представления атрибутов объектов (число, время, дата, текст и т.п.).

CDSSK.3: Структура справочника атрибутов.

Имя (URN) справочника атрибутов формируется в виде A_префикс-класса.

Элементы:

название атрибута;

формат представления значений атрибута (URN соответствующего элемента справочника объектов класса “Форматы данных”);

URN словаря значений атрибута (формируется в форме N_URN атрибута);

тип словаря значений атрибута (S — статический, D — динамический)³;

дополнительная информация (пояснительный текст).

Примеры:

A_UNFT.1: тип представления данных; ; N_A_UNFT.1; S; при вводе объекта выбирается из словаря

A_UNFT.2: вид формата; ; N_A_UNFT.2; S; выбирается из словаря

A_UNFT.3: обязательное (r) или факультативное (f) значение атрибута; ; N_A_UNFT.3; S;

A_UNFT.4: уникальное (u) или множественное (m) значение атрибута; ; N_A_UNFT.4; S;

A_UNPS.1: фамилия персоны; UNFT.i⁴; N_A_UNPS.1; D; фамилия выбирается из словаря, при отсутствии она вводится и проверяется на эквивалентность с другими написаниями.

A_UNPS.8: квалификация (ученая степень); UNFT.j⁵; N_A_UNPS.8; S; выбирается из словаря

A_UNBD.4: URL ресурса; UNFT.10⁶; N_A_UNBD.4; D;

CDSSK.4: Структура справочника универсальных связей.

Имя (URN) справочника: REUN.

Элементы:

название;

URN-значение словаря формата данных, определяющего форму представления данной связи;

описание связи.

Примеры:

REUN.1: Эквивалентность; N_A_UNFT.2.6. используется для обозначения идентичных атрибутов или связей (разные написания фамилий и имен, переводные версии одной публикации, разные наименования одной организации, синонимы терминов и т.п.).

REUN.3: Содержит; N_A_UNFT.2.6; сборник по отношению к его статьям, континент по отношению к расположенным на нем странам,

³ Статический словарь содержит “стандартные значения элементов” — например, названия научных степеней, рубрики классификационных систем и т.п.; динамический словарь наполняется значениями атрибутов по мере формирования контента пространства. Это — фамилии персон, наименования конкретных объектов и их свойства. Если рассматривать процедуру ввода данных в ЕЦПНЗ, то формировать статические словари могут пользователи со статусом “редактор”, а динамические — пользователи более низкого уровня со статусом “оператор”.

⁴ Этот элемент словаря форматов (i принимает при вводе реальных объектов конкретное значение) указывает, что значение атрибута может содержать только буквы, является обязательным и уникальным.

⁵ Этот элемент словаря форматов говорит о том, что значение атрибута факультативное, множественное, а тип словаря “S” указывает, что оно выбирается из ранее сформированного словаря.

⁶ Этот элемент словаря форматов содержит требования к форме представления URL, проверке достоверности ссылки и указывает, что элемент обязательный и уникальный.

организация по отношению к ее подразделениям и т.п.

CDSSK.5. Структура справочника квазиуниверсальных связей.

Имя (URN): **REQU**.

Элементы:

название;

префикс класса объекта;

необходимость словаря значений (Y/N);

URN словаря значений (если предыдущий атрибут = Y);

URN элемента словаря формата данных, определяющего форму представления данной связи;

описание связи.

Пример:

REQU.1: Начало временного интервала; **UNTC; N; ; N_A_UNFT.2.6;** значение определяется URN элемента словаря временных значений **UNTC**, на который будет дана ссылка в значении атрибута связи.

Здесь **UNTC** – префикс универсального класса “временные характеристики”, **N_A_UNFT.2.6** – URN элемента словаря значений “вид формата” (см. выше), описывающего структуру данного вида связи.

CDSSK.6: Структура справочника специфических связей.

Имя (URN) справочника: **RESP**.

Справочник имеет “шапку” из шести атрибутов, которая в случае составной связи дополняется блоками, содержащими по три атрибута, описывающими иерархию значений связи.

Элементы:

название связи;

префикс класса субъекта;

префикс класса объекта;

формат представления связи (URN значения соответствующего элемента словаря **N_A_UNFT.2**);

URN справочника атрибутов связи;

количество значений атрибутов, имеющих подчиненные связи следующего уровня (ноль или целое число n).

Если шестой элемент не равен нулю, то добавляется n блоков связи второго уровня:

URN элемента словаря значений атрибута, имеющего подчиненную связь;

URN справочника атрибутов подчиненной связи;

количество подчиненных связей следующего уровня (0–k).

Следующие n–1 блоков содержат информацию о связях второго уровня, а далее идут блоки, описывающие связи следующих уровней.

Пример:

RESP.3: идентификатор персоны в базе данных; **UNPS; UNBD; N_A_UNFT.2.7; A_RESP.3; 0**

Здесь **UNBD** – префикс универсального класса “базы данных”,

N_A_UNFT.2.7 – элемент словаря значений класса “форматы” (см. далее),

Справочник атрибутов связи **A_RESP.3** имеет два элемента. Первый из них определяет, с какой именно базой устанавливается связь (информация об этой базе должна быть описана как объект класса **UNBD**). Словарь **N_A_UNBD.1** содержит названия баз данных, откуда выбирается нужное, а при отсутствии информации о базе данных ее атрибуты вводятся в соответствующие словари по справочнику атрибутов. Второй элемент справочника определяет идентификатор конкретной персоны в этой базе данных, который записывается в словарь **N_A_RESP.3**, специфичный для данной связи. Оба словаря могут пополняться в процессе наполнения контента ЕЦПНЗ.

A_RESP.3.1: название базы данных; ; **N_A_UNBD.1; D;**

A_RESP.3.2: индекс персоны в базе данных; ; **N_A_RESP.3; D;**

Примером многоуровневой связи служит связь персоны с организацией (**RESP.4**), которая может принимать значения “сотрудник”, “спонсор”, “акционер” и т.п. Если персона является (являлась) сотрудником организации, то необходимо указать, в каких должностях; для каждой должности, в свою очередь, следует указать начало и окончание работы в этой должности. Структура этой связи записывается как элемент справочника **RESP.4**, но мы ее здесь не приво-

дим, поскольку это потребует большого количества определений справочников атрибутов и словарей их значений, что существенно перегрузит текст статьи.

CDSSK.7: Структура словаря значений атрибутов объектов и связей.

URN словаря формируется в форме N_URN атрибута. Словарь имеет один элемент — значение, URN формата которого указан в соответствующем элементе справочника атрибутов. Например:

$N_A_UNBD.1.2$: Большая российская энциклопедия.

$N_A_UNBD.4.2$: <https://bigenc.ru/>.

$N_A_UNFT.2.6$: простая связь URNc первого типа между объектами, атрибутами или значениями $O1$ и $O2$, имеющими соответственно $URNO1$ и $URNO2$, вида $\langle URNc \rangle : \langle URNO1 \rangle \langle URNO2 \rangle$, где URNc — URN конкретной связи.

$N_A_UNFT.2.7$: простая связь URNc второго типа между объектами $O1$ и $O2$, имеющими соответственно $URNO1$ и $URNO2$, принимающая некоторое значение, представленное в соответствующем словаре под именем URNd; формат связи имеет вид: $\langle URNc \rangle : \langle URNO1 \rangle \langle URNO2 \rangle = \langle URNd \rangle$.

CDSSK.8: Структура словарей связей.

Элементы словаря связи имеют в качестве URN имя справочника данной связи и отделенный от него точкой порядковый номер.

Каждый элемент словаря содержит информацию о связи пары URN конкретных объектов или URN значений атрибутов, представленную в формате, указанном в соответствующем справочнике связей.

Например, информация о том, что фамилия Савин (имеющая $URN=N_A_UNPS.1.3$) эквивалентна значению Savin (имеющему $URN=N_A_UNPS.1.53$), запишется (если предыдущие к элементам словаря связи “эквивалентно” до этого были заполнены) в виде элемента:

$REUN.1.k+1 : \langle N_A_UNPS.1.3 \rangle \langle N_A_UNPS.1.53 \rangle$

Информация о том, что в состав журнала, имеющего $URN=UNPB.i$, входит статья, имеющая $URN=UNPB.j$, представляется элементом словаря

$REUN.3.n : \langle UNPB.i \rangle \langle UNPB.j \rangle$.

Информация о том, что конкретная персона (совокупность информации о которой представлена в элементе словаря персон $UNPS.k$) имеет в РИНЦ идентификатор NNN, будет представлена в виде:

$RESP.3.m : \langle UNPS.k \rangle \langle UNBD.p \rangle = \langle N_A_RESP.3.q \rangle$,

где UNBD.p содержит информацию о РИНЦ, а $N_A_RESP.3.q$ — идентификатор NNN.

CDSSK.9: Структура словарей объектов.

Имя словаря совпадает с URN класса, к которому относится данный объект. Элемент словаря представляет собой перечень URN элементов словарей атрибутов и связей, относящихся к данному объекту. В частности, в нем должны присутствовать URN значений атрибутов, объявленных обязательными, что контролируется в процессе ввода данных по соответствующему элементу словаря форматов. Прежде чем привести пример словаря атрибутов, рассмотрим пример формализации описания специфической связи персоны с публикацией.

Каждый цифровой объект, принадлежащий ЕЦПНЗ, относится к одному и только одному классу и представляет собой совокупность значений характеризующих его атрибутов и связей с другими объектами, которая является одной из составляющих словаря объектов данного класса.

Предложенная структура описания контента ЕЦПНЗ позволяет добавлять в пространство новые подпространства, а в них — новые виды объектов, атрибуты и связи. Унификация и структуризация элементов существенно упрощают и ускоряют компьютерную обработку данных. Введение класса объектов “форматы” позволяет построить типовой набор алгоритмов формально-логического контроля данных и автоматически выбирать, какие из них в каком случае необходимы.

Составляющие всех динамических словарей формируются автоматически в процессе ввода данных в ЕЦПНЗ. Это может быть реализовано программным путем, если контент импортируется из внешней структуры — прикладная программа пакетного ввода данных обрабатывает справочники атрибутов и связей и записывает элементы в соответствующие словари. Если же объекты вводятся оператором, то работает диа-

логовая программа, которая на основе обработки справочников предлагает оператору сформировать значения атрибутов вводимого объекта и связей с другими объектами. По каждому атрибуту и связи оператор должен выбрать уже имеющиеся в ЕЦПНЗ значения или ввести новые. Контроль корректности ввода осуществляется автоматически с использованием информации о форматах данных, присутствующей в соответствующих справочниках.

Все значения атрибутов объектов хранятся в единственном экземпляре в отдельных словарях. При обработке запроса, включающего заданные значения тех или иных атрибутов и связей, на первом шаге поискового алгоритма заданные значения заменяются на их URN (или группу URN с учетом булевой логики запроса) и дальнейшая обработка происходит на уровне структурированных URN, что существенно ускоряет процесс поиска нужных объектов и навигации по связанным ресурсам. Результатом поиска являются элементы словарей объектов и связей, которые содержат URN элементов запроса. При визуализации результатов поиска происходит обратное преобразование URN в значащую информацию на основе справочников, имена которых жестко связаны со словарями.

В настоящее время ведутся работы по моделированию предложенной структуры онтологии ЕЦПНЗ на примере универсального подпространства. На текущий момент выделено 12 вспомогательных и 10 предметных классов объектов.

В качестве вспомогательных выделены следующие классы объектов:

- форматы данных;
- универсальные классификационные системы;
- научные направления;
- местоположение (пространственные характеристики);
- временные характеристики, даты;
- количественные характеристики;
- качественные характеристики;
- мировые константы;
- единицы измерения;
- языки;

- группы персон;
- коллекции;
- числовые значения.

К предметным классам УПП отнесены следующие:

- персоны;
- публикации;
- квалификационные работы;
- документы;
- музейные предметы;
- изображения и мультимедийные объекты;
- события/мероприятия;
- организации;
- политематические базы данных, каталоги ресурсов;
- награды, гранты.

В большинстве случаев из названия класса ясно, что собой представляют его объекты. Но такой класс, как “научные направления”, при наличии класса “классификационные системы” требует некоторого пояснения. В ЕЦПНЗ этот класс нужен для описания процессов, связанных с комплексными исследованиями, сформулированными в документах, связанных с планированием развития науки. Очевидно, что каждый объект этого класса связан с одним или несколькими “постановлениями”, относящимися к классу “документы”, специфической связью типа “впервые сформулировано”, а также с множеством объектов класса “универсальные классификационные системы” — конкретными разделами ГРНТИ, УДК, ВАК и др. Последнее необходимо для проведения комплексного анализа результатов исследований по тому или иному научному направлению.

Введен также класс “группы персон”, который необходим для указания, на какие категории пользователей ориентированы те или иные научные ресурсы (например, учебники для студентов по конкретной специальности, научно-популярные издания для школьников старших классов и т.п.).

Для каждого класса сформированы соответствующие справочники и элементы статических словарей. Определен ряд специфических связей

между объектами разных классов и значения этих связей.

Приведем несколько примеров таких связей.

Простая специфическая связь персоны с публикацией, которая имеет множественные атрибуты (одна персона может выступать в нескольких ролях):

- автор;
- редактор;
- составитель;
- автор перевода;
- художник;
- о нем;
- владелец авторских прав.

Простая специфическая связь персоны с музейным предметом, включающая множественные атрибуты:

- автор (изготовитель);
- автор сбора (для естественнонаучных коллекций);
- владелец;
- даритель.

Составная квазиуниверсальная связь публикации с классификационной системой — она состоит из двух составляющих: URN наименования классификационной системы и URN индекса, которому соответствует публикация.

Простая специфическая связь между научным открытием и публикацией с атрибутами:

- впервые опубликовано;
- рекомендованный учебник для вузов⁷;
- рекомендованное научно-популярное издание.

По мере разработки отдельных элементов они реализуются на примере ЭБ ННР, как модели фрагмента ЕЦПНЗ. В текущей версии ЭБ ННР (<http://e-heritage.ru>) в опции “Расширенный поиск” реализован поиск публикаций, персон, музейных объектов, отвечающих заданным значениям связей, приведенных выше.

⁷ Очевидно, что под понятием “рекомендованное” подразумевается, что издание признано таковым на основе серьезной научной экспертизы, что является одной из серьезных задач при отборе контента ЕЦПНЗ.

4. ЗАКЛЮЧЕНИЕ

Создание ЕЦПНЗ предполагает построение его онтологии, а также онтологий тематических подпространств, требует скоординированных усилий ведущих специалистов в области информационных технологий, представителей различных научных областей и держателей информационных ресурсов.

Исследования, связанные с формированием ЕЦПНЗ, проводятся в МСЦ РАН в рамках государственного задания по теме FNEF-2024-0014.

СПИСОК ЛИТЕРАТУРЫ

1. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестник Российской академии наук. 2019. Т. 89. № 7. С. 728–735
2. Васильева М.М. Формирование единого информационного пространства России в условиях глобализации // Вестник московского государственного лингвистического университета. 2013. № 684. С. 92–104.
3. Саенко И.Б. Подход к построению системы показателей качества единого информационного пространства // Естественные и математические науки в современном мире. 2014. № 14. С. 51–56.
4. Вершинин А.П. Единое российское электронное пространство знаний: вопросы права // Информационно-аналитический журнал “Университетская книга”, декабрь 2016 г. <http://www.unkniga.ru/biblioteki/bibdelo/6630-edinoe-rossiyskoe-elektronnnoe-prostranstvo-znaniy-voprosy-prava.html>
5. Single digital space of scientific knowledge: problems and solutions: collection of scientific papers. Moscow, Berlin: Directmedia Publishing, 2021. 502 p.
6. Большая российская энциклопедия. <https://bigenc.ru/>
7. Национальная электронная библиотека. <https://rusneb.ru/>
8. Российский индекс научного цитирования. https://elibrary.ru/project_risc.asp
9. Государственный каталог географических названий. <https://cgkipd.ru/science/names/reestry-gkgn.php>
10. ЭБ “Научное наследие России”. <http://e-heritage.ru>
11. Kalenov N.E., Sobolevskaya I.N., Sotnikov A.N., Common Digital Space of Scientific Knowledge Ontology // Lobachevskii Journal of Mathematics. 2023. Vol. 44. No. 7. P. 2733–2743.
12. W3C 2009. SKOS Simple Knowledge Organization System Reference. W3C Recommendation 18 August

2009. Available at <https://www.w3.org/TR/skos-reference/>
13. SKOS Simple Knowledge Organization System. <https://webarchive.library.unt.edu/web/20170125143526/http://www.w3.org/TR/skos-reference/#xl-Label>
 14. Zeng M.L., Mayr Ph. Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review // International Journal on Digital Libraries. 2018. <https://arxiv.org/pdf/1801.04479.pdf/>
 15. Pattuelli M.C., Provo A., Thorsen H. Ontology building for Linked Open Data: A pragmatic perspective. Journal of Library Metadata. 2015. Vol. 15. I. 3–4. P. 265–294.
 16. Resource Description Framework (RDF): Concepts and Abstract Syntax. <https://clck.ru/gwVBC>

COMMON DIGITAL SPACE OF SCIENTIFIC KNOWLEDGE AS AN INTEGRATOR OF POLYTHEMATIC INFORMATION RESOURCES

N. Kalenov^a, A. Sotnikov^a

^a*Joint SuperComputer Center of the Russian Academy of Sciences — Branch of Federal State Institution
“Scientific Research Institute for System Analysis of the Russian Academy of Sciences”.*

Presented by Academician of the RAS G. I. Savin

The goals, objectives and structure of the ontology of the Common Digital Space of Scientific Knowledge (CDSSK) are considered. The CDSSK is an integrated information structure that combines state scientific information systems presented on the Internet (the Great Russian Encyclopedia, the National Electronic Library, the State Catalog of Geographical Names, etc.) with industry information systems, databases and electronic libraries (“MathNet”, “Socionet”, “Scientific Heritage of Russia”, etc.). CDSSK can be considered as an information basis for solving artificial intelligence problems. The article presents the unified structure of the CDSSK ontology developed at the JSCC RAS and its modeling on the example of ten subject and eight auxiliary classes of objects of the CDSSK universal subspace.

Keywords: information resources, ontological approach, scientific knowledge space, structuring, content, semantic Web