

Диалоговый программный комплекс для формирования онтологии Единого цифрового пространства научных знаний

С.А. Власова^{1,2}✉, Н.Е. Каленов^{1,2}

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

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

Ссылка для цитирования

Власова С.А., Каленов Н.Е. Диалоговый программный комплекс для формирования онтологии Единого цифрового пространства научных знаний // Программные продукты и системы. 2024. Т. 37. № 4. С. 514–523. doi: 10.15827/0236-235X.148.514-523

Информация о статье

Группа специальностей ВАК: 2.3.5

Поступила в редакцию: 25.07.2024

После доработки: 23.08.2024

Принята к публикации: 27.08.2024

Аннотация. В статье рассматривается программный комплекс, формирующий онтологию Единого цифрового пространства научных знаний (ЕЦПНЗ). Онтология включает совокупность справочников, описывающих подпространства, классы объектов, атрибуты объектов и связи различных типов и видов. Каждый справочник атрибутов объектов содержит наименование атрибута, информацию о его формате и статусе, а также имя словаря значений и его тип. Справочники связей содержат необходимую информацию о связях пар объектов или значений атрибутов. Онтология ЕЦПНЗ включает именованные связи трех типов – универсальные, квазиуниверсальные и специфические. Конкретная связь любого типа может быть простой или составной. Представленный в статье программный комплекс реализует сервисы, связанные с формированием справочников подпространств, классов, атрибутов объектов и связей всех типов ЕЦПНЗ. Комплекс ориентирован на администраторов подпространств ЕЦПНЗ и позволяет в режиме диалога создавать структурные элементы ЕЦПНЗ. Эти элементы необходимы для формирования контента пространства с автоматическим определением списка атрибутов объектов, форматов их ввода и необходимых средств формально-логического контроля. При работе комплекса обеспечивается контроль вводимой информации, исключая дублирование элементов или некорректный ввод данных, противоречащий декларированной классом «форматы» структуре. Программный комплекс состоит из двух модулей – формирования новых элементов и редактирования уже имеющихся. Модуль редактирования позволяет удалять элементы онтологии без нарушения ее целостности. В результате работы комплекса строится структура, являющаяся основой для унифицированных программ формирования контента ЕЦПНЗ.

Ключевые слова: программный комплекс, цифровое пространство научных знаний, онтология, классы объектов, атрибуты, структуризация, связанные данные

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

Введение. Структурированная информационная среда *Единое цифровое пространство научных знаний* (ЕЦПНЗ) объединяет сведения о различных объектах, связанных с наукой, в качестве которых могут выступать цифровая копия физической сущности (например, книги, музейного предмета, архивного документа и т.п.), база данных, информация об ученом, научном мероприятии, научном факте и т.п. [1, 2].

Учитывая основные тренды современного развития информатики, связанные с автоматизацией обработки научной информации и применением искусственного интеллекта во многих областях науки, ЕЦПНЗ необходимо строить на принципах Semantic Web [3, 4], а при его разработке использовать онтологический подход [5].

Для описания онтологий и проектирования на их основе информационных систем существует много языков, основой которых является принятый в качестве стандарта OWL,

базирующийся на представлении данных в структурах RDF (<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>) и RDFS (<https://www.w3.org/TR/rdf-schema/>). На базе этих стандартов консорциумом W3 разработана модель организации знаний SKOS (*Simple Knowledge Organization System*), унифицирующая структуру представления тезаурусов, относящихся к различным областям знаний (<https://www.w3.org/TR/skos-reference/>).

В Сети представлено значительное количество публикаций, посвященных разработкам на основе OWL-онтологий, описывающих отдельные направления науки и техники как за рубежом (например, [6–8]), так и в России (например, [9–11]).

Однако, как показывает анализ, использование OWL затруднительно при описании объектов, обладающих многоуровневыми атрибутами (свойствами – в терминологии OWL), такими как численность населения страны или города

по годам, расход воды в водопадах по месяцам, работа персоны в той или иной организации в разных должностях в разное время и т.п. Эта проблема и сложности ее преодоления подчеркнуты, в частности, в [12]. Отражая динамику развития различных областей науки, в том числе в историческом плане, онтология ЕЦПНЗ должна учитывать такие аспекты, как переименование населенных пунктов, изменение принадлежности территорий тем или иным государствам, различные обозначения единиц измерения и т.п.

Необходимо отметить, что в информатике существует ряд трактовок термина «онтология» применительно к различным задачам. В данной работе под онтологией будем понимать структурированные данные, отражающие информацию об элементах ЕЦПНЗ и их связях.

Элементы онтологии ЕЦПНЗ

ЕЦПНЗ представляет собой совокупность тематических *подпространств* (ПП), связанных единой онтологией – унифицированными правилами отражения информации об объектах и их связях, а также собственно связями между объектами различных ПП.

Структура онтологии ЕЦПНЗ описана в [13–15] и представлена в виде элементов справочника CDSSK на сайте ЕЦПНЗ (<http://dirsmc.ru/cdssk/structure/>). Чтобы облегчить понимание работы программного комплекса, приведем основные определения, относящиеся к структуре онтологии ЕЦПНЗ.

Объект ЕЦПНЗ характеризуется фиксированным набором атрибутов (свойств), каждый из которых принимает определенное значение (или ряд значений) для конкретного объекта. Перечень атрибутов объектов определяется задачами ЕЦПНЗ, к решению которых имеет отношение данный объект.

Объекты ПП с одинаковыми наборами атрибутов объединяются в классы.

Между парами любых объектов и значений атрибутов могут быть установлены связи, каждая из которых характеризуется видом, типом и (факультативно) значением.

Структуры классов, атрибутов объектов и связей определяются унифицированными справочниками и соответствующими объектами класса «форматы» универсального ПП.

Каждый элемент ЕЦПНЗ имеет свое уникальное имя (URN). Упростить обработку данных для формирования URN позволяют следующие введенные мнемонические правила.

Каждому формируемому ПП присваивается префикс, который в общем случае может состоять из фиксированного числа символов, определяемого в справочнике ЕЦПНЗ. В данной версии префикс ПП состоит из двух буквенно-цифровых символов, включающих стандартную латиницу и кириллицу.

URN класса объектов состоит из четырех буквенно-цифровых символов: первые два являются префиксом ПП, к которому принадлежит данный класс, два других – префиксом класса. Классы с одним и тем же префиксом, обозначающим группу описываемых сущностей, могут принадлежать различным ПП и иметь разный, но пересекающийся набор атрибутов. Например, персоны, включаемые в универсальное ПП, имеют более широкий набор атрибутов, чем персоны, включаемые в тематическое ПП; объекты класса «местоположение» универсального ПП имеют существенно меньше атрибутов, чем те же объекты, описываемые в ПП, относящихся к географии, и т.п. Эквивалентность отдельных объектов из разных ПП устанавливается с помощью соответствующей связи. В качестве URN класса не используются буквосочетания REUN, REQU, RESP, зарезервированные для обозначения универсальных, квазиуниверсальных и специфических связей.

Справочники ЕЦПНЗ представляют собой кортежи, их размерность определяется видом элемента, который они описывают (ПП, класс, атрибут, связи).

Значения атрибутов объектов и связей хранятся в словарях одного из двух видов – статических или динамических. Статические словари (в справочниках они обозначены буквой S) содержат стандартизированные значения атрибутов (перечень ученых степеней и званий персон, классификационные индексы, виды отношений между объектами различных классов и т.п.). Их элементы формируются либо при первоначальной загрузке системы в пакетном режиме, либо в процессе формирования справочников разработанным программным комплексом. Словари значений второго типа формируются в процессе наполнения контента ЕЦПНЗ (в справочниках они обозначены буквой D).

Предлагаемый программный комплекс предназначен для формирования справочников элементов ЕЦПНЗ всех видов и для наполнения статических словарей значений атрибутов объектов и связей. Он ориентирован на администратора, формирующего и редактирующего

структуру определенного ПП с его классами и связями объектов.

Результатом работы комплекса является система связанных справочников и статических словарей, URN которых построены по установленным правилам. Элементы справочников и статических словарей динамически формируются в виде реляционной базы данных, а также текстовых строк заданной структуры [14].

Одной из основных составляющих справочников ЕЦПНЗ является указание на формат представления элемента, к которому относится справочник. Каждый элемент формата представляет собой объект вспомогательного класса универсального ПП «форматы», имеющий 6 атрибутов [16] – «тип представления данных», «вид формата», «признак обязательности», «признак уникальности», «ограничения по структуре», «описание». Значения атрибутов формируются данным программным комплексом и хранятся в соответствующих статических словарях, а их комбинации, характеризующие правила представления конкретных атрибутов и связей, – в словаре форматов. URN соответствующих элементов словаря форматов являются обязательными составляющими всех справочников атрибутов объектов и связей.

Классификация связей ЕЦПНЗ

Для упрощения алгоритмов формирования справочников связей и сокращения диалога с администратором введена классификация связей. Каждая связь определяется типом и видом.

Выделены пять типов связей:

- универсальная связь между объектами;
- универсальная связь между значениями атрибутов;
- универсальная связь между значениями конкретного объекта;
- квазиуниверсальная связь между объектами;
- специфическая связь между объектами.

Связь определенного типа может относиться к одному из следующих видов:

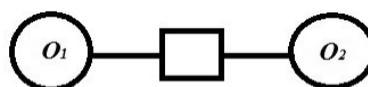
- простая связь, представляющая собой аналог триплетов RDF и содержащая указания на связи «субъект» и «объект» (здесь и далее приводимые в кавычках понятия «субъект» и «объект» понимаются в терминах триплета RDF, на рисунках они обозначены как O_1 и O_2 соответственно) (рис. 1);

- составная связь первого рода, содержащая (дополнительно к указаниям «субъект» и «объ-

ект») URN значений атрибутов связи, приведенных в соответствующих словарях, определяемых справочником атрибутов связи (рис. 2);

- составная связь второго рода, содержащая неразветвленную иерархическую цепочку значений атрибутов, когда все значения атрибута данного уровня имеют одни и те же подчиненные атрибуты (рис. 3);

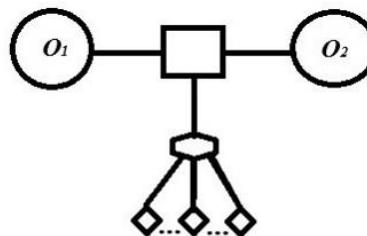
- составная связь третьего рода – разветвленная иерархическая цепочка, в которой каждый атрибут верхнего уровня имеет свои атрибуты следующего уровня (рис. 4).



○ - объект
□ - связь

Рис. 1. Простая связь

Fig. 1. Simple relation



⬡ атрибут связи
◇ значение атрибута

Рис. 2. Составная связь первого рода

Fig. 2. Composite relation of the first kind

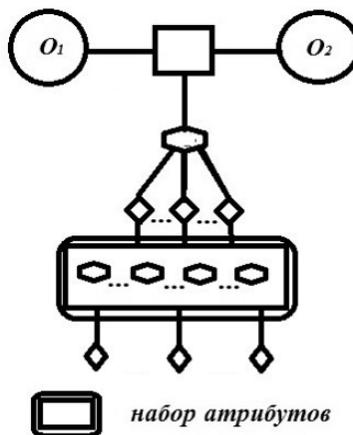
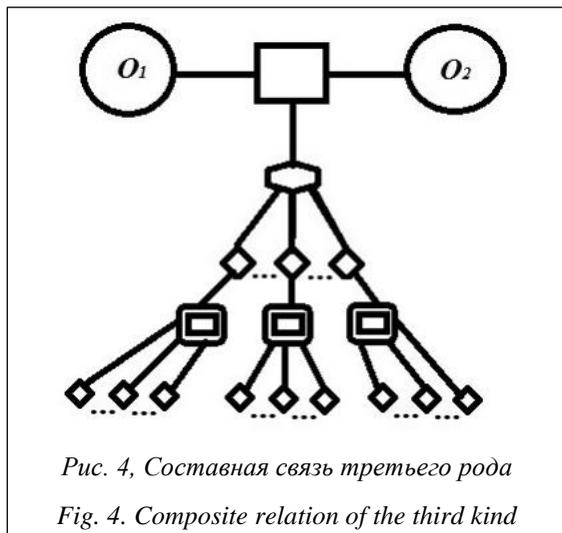


Рис. 3. Составная связь второго рода

Fig. 3. Composite relation of the second kind



Типы атрибутов объектов ЕЦПНЗ

Атрибуты могут иметь простые или составные значения, а также принимать значения связей.

Простое значение атрибута определяется элементом словаря значений атрибута, заданного в его справочнике.

Составное значение атрибута может состоять из нескольких значений одного уровня (одноуровневый составной атрибут) или из нескольких иерархических значений (многоуровневый составной атрибут). К одноуровневым составным атрибутам относятся, например, количественные и временные характеристики объекта – высота, длина, скорость и т.п. Значение такого атрибута состоит из единицы измерения и ее числового выражения. Примером многоуровневого составного атрибута может служить расход воды в водопаде или реке (количественная характеристика в зависимости от времени года), численность населения административной единицы с указанием года, скорость звука в различных средах и т.п.

Значение атрибута как связи может вводиться для удобства пользовательского интерфейса при диалоговом вводе данных. Так, при формировании метаданных доклада на конференции удобнее запросить ввод атрибута «автор доклада», чем запрашивать связь доклада с персоной.

Вид атрибута определяется его справочником, содержащим указание на соответствующий элемент формата, имя словаря значений атрибута и имя справочника подчиненного атрибута (при его наличии). Для простого атрибута в справочнике будет указано имя словаря значений атрибута, для составного атрибута –

имя словаря соответствующего объекта, для атрибута как связи – имя словаря соответствующей связи.

Для автоматического формирования системы связанных справочников и статических словарей ЕЦПНЗ разработан специальный диалоговый программный комплекс ФОНТ ЕЦПНЗ. Он создан на основе технологии Microsoft ASP.NET на платформе Microsoft .NET Framework в среде разработки Microsoft Visual Studio.

Структура программного комплекса ФОНТ ЕЦПНЗ

Комплекс включает модули создания элементов ЕЦПНЗ (ПП, классов объектов, атрибутов объектов и связей) и редактирования их справочников и статических словарей значений атрибутов.

В модуле создания ПП пользователю предоставляется список наименований имеющихся ПП (рис. 5). Наименования являются активными ссылками, обеспечивающими получение данных о ПП с возможностью их редактирования. Для регистрации нового ПП необходимо ввести наименование ПП, префикс (2 символа), описание ПП (рис. 5).

Модуль формирования классов предлагает список наименований имеющихся ПП, из него нужно выбрать то, в котором будет создан класс. Для регистрации нового класса нужно ввести его наименование и префикс (2 символа). Автоматически будет сформирован URN справочника атрибутов для данного класса, который при необходимости можно изменить.

Для создания атрибута из списка наименований имеющихся ПП выбирается нужное, затем из списка классов данного ПП выбирается класс, в котором будет создан атрибут. На рисунке 6 показана страница создания справочника нового атрибута в классе «административные единицы» ПП «Социально-экономическая география». В правой части экрана выдается список наименований имеющихся атрибутов данного класса. Наименования являются активными ссылками, обеспечивающими получение данных справочника атрибута и словаря его значений.

Просмотр данных атрибута «вид объекта»:

Наименование атрибута

вид объекта

URN атрибута **A_39AU.2**

Формат представления значений атрибута

UNFT . 42

URN словаря значений атрибута

N_A_39AU.2

Тип словаря значений **S**

Словарь значений атрибута

часть света

страна

регион / область / штат / край

город / поселок / деревня

Создание нового атрибута начинается с ввода его названия в левой части экрана (рис. 6).

Далее определяется формат атрибута. Из списка имеющихся типов формата выбирается нужный тип, например, выбор типа формата для атрибута «численность населения»:

Подпространство: **Социально-экономическая география**

Класс: **административные единицы**

Наименование атрибута: **численность населения**

Тип формата

- текст
- изображение
- видео
- аудио
- любое число
- ссылка на внешний ресурс
- дата в формате гггг[.мм[.дд]]
- количественная характеристика**
- временная характеристика

Затем из списка форматов указанного типа выбирается формат, например, для атрибута «численность населения»:

Рис. 5. Регистрация нового ПП

Fig. 5. Registration of a new subspace

Рис. 6. Создание справочника атрибута класса «административные единицы»

Fig. 6. Creating a directory of the class attribute “administrative units”

Выберите формат

- Количественная характеристика атрибут обязательный, значение повторяется
- **Количественная характеристика атрибут факультативный, значение повторяющееся**
- Факультативная уникальная количественная характеристика
- Другой вид формата

Если нужного формата в списке нет, то система предоставляет возможность ввода нового формата.

После выбора формата автоматически формируется URN словаря значений атрибута. В данном случае выбран URN словаря значений атрибута «количественная характеристика» (рис. 7). При вводе простого формата по умолчанию формируется имя словаря его значений в форме N_<URN справочника атрибута>, которое можно при необходимости изменить. В случае формирования атрибута «как связь» или составного атрибута (тип формата атрибута «количественная характеристика», «временная характеристика» и т.п.) URN словаря значений атрибута изменить нельзя. Для завершения формирования справочника атрибута нужно указать тип словаря значений (S – статический или D – динамический) и при необходимости ввести дополнительную информацию (рис. 7). Для словарей типа S предоставляется возможность ввода значений для данного атрибута.

После формирования справочника нового атрибута можно перейти к формированию атрибута второго уровня. Так, например, для атрибута «численность населения» можно сфор-

Создание справочника атрибутов

Подпространство: Социально-экономическая география Класс: административные единицы

Наименование атрибута: численность населения
Формат представления значений атрибута: UNFT.46

URN словаря значений атрибута: UNQC

Тип словаря значений (S - статический, D - динамический)

S
 D

Дополнительная информация

Рис. 7. Создание справочника для атрибута «численность населения»

Fig. 7. Creating a directory for "population size" attribute

мировать подчиненный атрибут «дата указания количества населения» с форматом «дата в формате гггг[.мм[.дд]], атрибут необязательный, значение уникальное».

В программном комплексе ФОНТ ЕЦПНЗ модули создания ПП, классов и атрибутов связаны между собой удобной навигацией. После регистрации нового ПП можно сразу перейти к созданию классов для него. После регистрации класса перейти к созданию справочников атрибутов данного класса. При создании нового атрибута следует сразу ввести его значения в соответствующий статический словарь.

Программный комплекс предусматривает возможность редактирования наименований и описаний ПП и классов. Специальный модуль редактирования атрибутов позволяет изменять наименование и описание атрибута, URN и тип словаря значений (если атрибут еще нигде не использовался), добавлять и редактировать значения словаря атрибута (<http://www.swsys.ru/uploaded/image/2024-4/4.jpg>).

Модуль создания связей

В данном модуле процедура формирования новой связи определяется ее форматом. При создании любой связи необходимо вводить ее наименование и описание. Формирование квазиуниверсальной связи требует определения объекта связи, а специфической – и субъекта, и объекта. При создании связи с форматами «универсальная составная связь первого рода между объектами», «универсальная составная связь первого рода между значениями атрибутов», «квазиуниверсальная составная связь первого рода», «специфическая составная связь первого рода» формируются справочник атрибута связи и словарь значений атрибута.

В связях с форматами «универсальная составная связь второго рода между объектами», «универсальная составная связь второго рода между значениями атрибутов конкретного объекта», «специфическая составная связь второго рода» формируется справочник атрибута связи. Если тип словаря значений атрибута статический, то вводятся значения атрибута. Также формируются справочники атрибутов второго уровня (относятся ко всем значениям атрибута первого уровня) и вводятся их значения (в случае статического типа словаря значений).

При создании связи с форматом «специфическая составная связь третьего рода» формируются справочник атрибута связи и словарь значений атрибута. Запрашивается количество

подчиненных связей. Выбирается значение атрибута, для него формируются справочники подчиненных атрибутов и вводятся значения для этих атрибутов (в случае статического типа словаря значений). Последняя процедура повторяется для всех значений атрибута первого уровня, у которых должны быть подчиненные атрибуты.

Работу модуля формирования связей рассмотрим на примере создания связи персоны с публикацией. После выбора типа связи «Специфическая связь» предоставляется список имеющихся специфических связей и предлагается ввести наименование новой связи и ее описание (<http://www.swsys.ru/uploaded/image/2024-4/5.jpg>). Затем формируются префикс субъекта связи (выбирается «Универсальное пространство», класс «Персоны»), префикс объекта связи (выбирается «Универсальное пространство», класс «Публикации») и выбирается формат связи «специфическая составная связь первого рода»:

Наименование связи: **связь персоны с публикацией**

Префикс класса субъекта: **UNPS**

Префикс класса объекта: **UNPB**

Формат представления связи

- o специфическая простая связь
- o **специфическая составная связь первого рода**
- o специфическая составная связь второго рода
- o специфическая составная связь третьего рода

Далее для формирования справочника атрибута вводится его наименование (в данном

примере – «роль персоны в создании публикации»), выбирается формат атрибута «текст, только буквы, атрибут обязательный, значение повторяющееся», указывается тип словаря значений атрибута «S – статический» (рис. 8). После ввода значений словаря данного атрибута (например, «автор», «редактор», «составитель») формирование связи будет полностью завершено. Представим все данные сформированной связи:

Наименование связи **связь персоны с публикацией**

URN связи **RESP.5**

Префикс класса субъекта **UNPS**

Префикс класса объекта **UNPB**

Формат представления связи **UNFT.53 специфическая составная связь первого рода**

URN справочника атрибута связи **A_RESP.5**

Наименование атрибута **роль персоны в создании публикации**

Формат значений атрибута **UNFT.9**

Текст, только буквы, атрибут обязательный, значение повторяющееся

URN справочника атрибута **A_RESP.5.1**

URN справочника значений атрибута **N_A_RESP.5.1**

Тип словаря значений **S**

Словарь значений атрибута

автор

редактор

составитель

автор перевода

художник

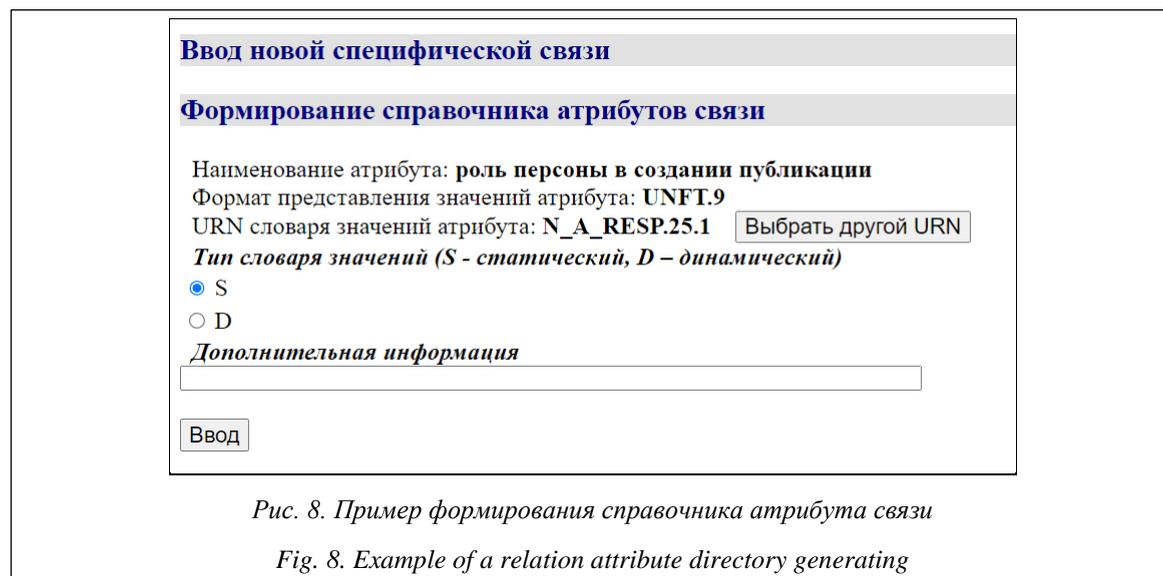


Рис. 8. Пример формирования справочника атрибута связи

Fig. 8. Example of a relation attribute directory generating

Специальный модуль редактирования связей позволяет вносить изменения в атрибуты связей и в словари их значений.

Заключение

Комплекс ФОНТ ЕЦПНЗ находится в стадии опытной эксплуатации и используется для моделирования процессов формирования различных элементов ЕЦПНЗ и связей между разнородными объектами, отражающими научные знания. В частности, сформирована онтология и загружены атрибуты и связи объектов, им-

портированных из БД публикаций сотрудников МСЦ РАН, в стадии загрузки находится одна из коллекций музейных объектов. В дальнейшем предполагается реализовать многоплатформенную версию комплекса. Сформированные с помощью комплекса варианты онтологии отдельных ПП являются базой для реализации унифицированных программных средств ЕЦПНЗ, обеспечивающих пакетную и диалоговую технологии наполнения пространства разнородными реальными научными данными, что составляет одну из очередных задач в области формирования ЕЦПНЗ.

Список литературы

1. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестн. РАН. 2019. Т. 89. № 7. С. 728–735. doi: 10.31857/S0869-5873897728-735.
2. Савин Г.И. Единое цифровое пространство научных знаний: цели и задачи // Информационные ресурсы России. 2020. № 5. С. 3–5. doi: 10.51218/0204-3653-2020-5-3-5.
3. Хорошевский В.Ф. Пространства знаний в сети Интернет и Semantic WEB (часть 1) // ИИиПР. 2008. № 1. С. 80–97.
4. Хорошевский В.Ф. Пространства знаний в сети Интернет и Semantic WEB (часть 2) // ИИиПР. 2009. № 4. С. 15–36.
5. Каленов Н.Е., Серебряков В.А. Об онтологии Единого цифрового пространства научных знаний // Информационные ресурсы России. 2020. № 5. С. 10–12. doi: 10.51218/0204-3653-2020-5-10-12.
6. Alromema W., Alahmadi A.H. Ontology building for patient bioinformatics of the smart card domain using OWL. *IJCC*, 2022, vol. 11, no. 4, pp. 316–329. doi: 10.1504/IJCC.2022.124796.
7. Luo Yu., Xiong X., Jin Sh., Liu Z. VIM: Research on OWL-based vocabulary ontology construction method for units of measurement. *Electronics*, 2023, vol. 12, no. 18, art. 3783. doi: 10.3390/electronics12183783.
8. Kosińska J., Brotoń G., Tobiasz M. Knowledge representation of the state of a cloud-native application. *STTT*, 2024, vol. 26, no. 1, pp. 21–32. doi: 10.1007/s10009-023-00705-2.
9. Загорюлько Ю.А., Боровикова О.И. Использование системы разнородных паттернов онтологического проектирования для разработки онтологий научных предметных областей // Программирование. 2020. № 4. С. 27–35. doi: 10.31857/S0132347420040068.
10. Загорюлько Ю.А., Сидорова Е.А., Загорюлько Г.Б., Ахмадеева И.Р., Серый А.С. Автоматизация разработки онтологий научных предметных областей на основе паттернов онтологического проектирования // Онтология проектирования. 2021. Т. 11. № 4. С. 500–520. doi: 10.18287/2223-9537-2021-11-4-500-520.
11. Митрофанова О.А., Митрофанов Е.П., Буре Н.А. Применение онтологического подхода к проектированию геопространственной базы опытных данных для информационного обеспечения исследований в точном земледелии // Вестн. СПбГУ. Прикладная математика. Информатика. Процессы управления. 2022. Т. 18. № 2. С. 253–262. doi: 10.21638/11701/spbu10.2022.206.
12. Марчук А.Г., Марчук П.А. Базовая онтология неспецифических сущностей BONE и ее использование для построения информационных систем // Вестн. СибГУТИ. 2014. № 4. С. 118–128.
13. Каленов Н.Е., Сотников А.Н. Структура онтологии единого цифрового пространства научных знаний // НТИ. Сер. 2. 2023. № 7. С. 20–26. doi: 10.36535/0548-0027-2023-07-3.
14. Kalenov N.E., Sobolevskaya I.N., Sotnikov A.N. Common digital space of scientific knowledge ontology structuring. *Lobachevskii J. of Math.*, 2023, vol. 44, pp. 2733–2743. doi: 10.1134/S1995080223070235.
15. Kalenov N.E., Sotnikov A.N. Common digital space of scientific knowledge as an integrator of polythematic information resources. *Doklady Math.*, 2024, vol. 109, pp. 93–99. doi: 10.1134/S106456242470176X.
16. Власова С.А., Каленов Н.Е., Сотников А.Н. Онтология вспомогательных и политематических предметных классов единого цифрового пространства научных знаний // Электронные библиотеки. 2024. Т. 27. № 1. С. 22–42. doi: 10.26907/1562-5419-2024-27-1-22-42.

Dialogue software complex for forming Common Digital Space of Scientific Knowledge ontologySvetlana A. Vlasova ^{1,2}✉, Nikolay E. Kalenov ^{1,2}¹ Joint Supercomputer Center of RAS, Moscow, 119991, Russian Federation² National Research Centre “Kurchatov Institute”,
Moscow, 123182, Russian Federation**For citation**Vlasova, S.A., Kalenov, N.E. (2024) ‘Dialogue software complex for forming Common Digital Space of Scientific Knowledge ontology’, *Software & Systems*, 37(4), pp. 514–523 (in Russ.). doi: 10.15827/0236-235X.148.514-523**Article info**

Received: 25.07.2024

After revision: 23.08.2024

Accepted: 27.08.2024

Abstract. The paper considers a software package that forms the ontology of the Common Digital Space of Scientific Knowledge (CDSSK). The ontology includes a set of directories describing subspaces, classes of objects, object attributes and relations of various types and kinds. Each object attribute directory contains an attribute name, information about its format and status (obligatory or optional, unique or multiple), as well as the name and type of the value dictionary. Relationship directories contain necessary information about the connections between object pairs or attribute values. The CDSSK ontology includes named connections of three types: universal, quasi-universal, and specific. A specific connection of any type can be simple or composite. The paper presents the software package that implements services related to the formation of reference books of subspaces, classes, attributes of objects and connections of all CDSSK types. The software package is for the administrators of CDSSK subspaces; it allows creating CDSSK structural elements in the dialog mode. These elements are necessary for forming the space content with automatic definition of the list of object attributes, their input formats and the necessary means of formal and logical control. The complex operation enables the input information control that excludes elements duplication or incorrect data input, which contradicts the structure declared by the class “formats”. The software package consists of two modules: forming new elements and editing existing ones. The editing module also allows deleting ontology elements without violating its integrity. The complex work enables building a structure that is a basis for unified programs of CDSSK content formation.

Keywords: software package, digital space of scientific knowledge, ontology, object classes, attributes, structuring, linked data

Acknowledgements. The paper was carried out under the government assignment, project No. FNEF-2024-0014

References

1. Antopolskii, A.B., Kalenov, N.E., Serebryakov, V.A., Sotnikov, A.N. (2019) ‘Common digital space of scientific knowledge’, *Bull. of RAS*, 89(7), pp. 728–735 (in Russ.). doi: 10.31857/S0869-5873897728-735.
2. Savin, G.I. (2020) ‘Common digital space of scientific knowledge: Goals and tasks’, *Inform. Resources of Russia*, (5), pp. 3–5 (in Russ.). doi: 10.51218/0204-3653-2020-5-3-5.
3. Khoroshevskiy, V.F. (2008) ‘Knowledge spaces on the Internet and Semantic WEB (part 1)’, *Artificial Intelligence and Decision Making*, (1), pp. 80–97 (in Russ.).
4. Khoroshevskiy, V.F. (2009) ‘Knowledge spaces on the Internet and Semantic WEB (part 2)’, *Artificial Intelligence and Decision Making*, (4), pp. 15–36 (in Russ.).
5. Kalenov, N.E., Serebryakov, V.A. (2020) ‘Ontology of the common digital space of scientific knowledge’, *Inform. Resources of Russia*, (5), pp. 10–12 (in Russ.). doi: 10.51218/0204-3653-2020-5-10-12.
6. Alromema, W., Alahmadi, A.H. (2022) ‘Ontology building for patient bioinformatics of the smart card domain using OWL’, *IJCC*, 11(4), pp. 316–329. doi: 10.1504/IJCC.2022.124796.
7. Luo, Yu., Xiong, X., Jin, Sh., Liu, Z. (2023) ‘VIM: Research on OWL-based vocabulary ontology construction method for units of measurement’, *Electronics*, 12(18), art. 3783. doi: 10.3390/electronics12183783.
8. Kosińska, J., Brotoń, G., Tobiasz, M. (2024) ‘Knowledge representation of the state of a cloud-native application’, *STTT*, 26(1), pp. 21–32. doi: 10.1007/s10009-023-00705-2.
9. Zagorulko, Yu.A., Borovikova, O.I. (2020) ‘Using a system of heterogeneous ontology design patterns to develop ontologies of scientific subject domains’, *Programming and Computer Software*, 46(4), pp. 273–280.
10. Zagorulko, Yu.A., Sidorova, E.A., Zagorulko, G.B., Akhmadeeva, I.R., Sery, A.S. (2021) ‘Automation of the development of ontologies of scientific subject domains based on ontology design patterns’, *Ontology of Designing*, 11(4), pp. 500–520 (in Russ.). doi: 10.18287/2223-9537-2021-11-4-500-520.
11. Mitrofanova, O.A., Mitrofanov, E.P., Bure, N.A. (2022) ‘Ontological approach application to the design of a geo-spatial experimental database for information support of research in precision agriculture’, *Vestn. of SPbSU. Applied Math. Comput. Sci. Control Processes*, 18(2), pp. 253–262 (in Russ.). doi: 10.21638/11701/spbu10.2022.206.
12. Marchuk, A.G., Marchuk, P.A. (2014) ‘Basic ontology of non-specific entities BONE and its use for building information systems’, *The Herald of the SibSUTIS*, (4), pp. 118–128 (in Russ.).
13. Kalenov, N.E., Sotnikov, A.N. (2023) ‘Structure of the common digital space of scientific knowledge ontology’, *STI. Ser. 2*, (7), pp. 20–26 (in Russ.). doi: 10.36535/0548-0027-2023-07-3.

14. Kalenov, N.E., Sobolevskaya, I.N., Sotnikov, A.N. (2023) 'Common digital space of scientific knowledge ontology structurization', *Lobachevskii J. of Math.*, 44, pp. 2733–2743. doi: 10.1134/S1995080223070235.

15. Kalenov, N.E., Sotnikov, A.N. (2024) 'Common digital space of scientific knowledge as an integrator of polythematic information resources', *Doklady Math.*, 109, pp. 93–99. doi: 10.1134/S106456242470176X.

16. Vlasova, S.A., Kalenov, N.E., Sotnikov, A.N. (2024) 'Ontology of the universal subspace of common digital space of scientific knowledge', *Russ. Digital Libraries J.*, 27(1), pp. 22–42 (in Russ.).

Авторы

Власова Светлана Александровна^{1,2}, к.т.н.,
вед. науч. сотр., vlas.svetlana2013@yandex.ru
Каленов Николай Евгеньевич^{1,2}, д.т.н.,
профессор, гл. науч. сотр., nekalenov@mail.ru

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

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

Authors

Svetlana A. Vlasova^{1,2}, Cand. of Sci. (Engineering),
Leading Researcher, vlas.svetlana2013@yandex.ru
Nikolay E. Kalenov^{1,2}, Dr.Sci. (Engineering)
Professor, Chief Researcher, nekalenov@mail.ru

¹ Joint Supercomputer Center of RAS,
Moscow, 119334, Russian Federation

² National Research Centre “Kurchatov Institute”,
Moscow, 123182, Russian Federation