

Информационные системы и технологии

УДК 004.82

DOI:10.25729/ESI.2025.38.2.015

Модель конструктора информационных порталов на основе онтологий предметной области и ресурсов

Лисин Владислав Александрович¹, Сидорова Елена Анатольевна²

¹Новосибирский государственный университет,
Новосибирск, Россия, *vladlisin2@gmail.com*

²Институт систем информатики им. А.П. Ершова СО РАН,
Новосибирск, Россия, *lsidorova@iis.nsk.su*

Аннотация. в статье представлен подход к разработке конструктора информационных порталов знаний на основе онтологического описания информации. Используется методология построения порталов знаний на основе базовых онтологий и паттернов онтологического проектирования. Конструктор порталов знаний должен обеспечивать end-to-end технологию создания и наполнения специализированного интернет-ресурса и предоставлять ряд интеллектуальных сервисов, таких, как редактор онтологий для моделирования предметной области и ресурсов, инструмент для семантической разметки текстов на основе онтологии и настраиваемый интерфейс. Особенности предлагаемого подхода являются: а) использование базовых онтологий научной предметной области и ресурсов, как фундамент для разработки специализированных онтологий, б) использование онтологий для реализации ключевой функциональности программного обеспечения, связанного с настройкой визуализации данных, навигации и интеллектуального поиска на целевом ресурсе, в) ориентированность платформы на специалистов, не имеющих опыта в программировании и веб-разработке. В статье предложены формальная модель конструктора информационных порталов и модель информационного ресурса, создаваемого с помощью данного конструктора, описаны важные технические решения для поддержки end-to-end технологии. Созданные базовые компоненты технологии продемонстрированы на примере создания двух простых информационных систем.

Ключевые слова: редактор онтологий, онтология ресурсов, редактор страниц, портал знаний, мультимедиа

Цитирование: Лисин В.А. Модель конструктора информационных порталов на основе онтологий предметной области и ресурсов / В.А. Лисин, Е.А. Сидорова // Информационные и математические технологии в науке и управлении, 2025. – № 2 (38). – С. 182-198. – DOI:10.25729/ESI.2025.38.2.015.

Введение. Доступ к качественным и достоверным данным, возможность их изучения особенно важны в условиях перенасыщенности информацией. Для этих целей создаются специализированные распределенные системы, которые способствуют углубленному изучению предметной области. Одним из методов разработки таких систем является применение онтологий – формального описания предметной области (PrO), в качестве метакаркаса для систематизации знаний, хранения данных и поддержки аналитических и интеллектуальных сервисов.

В научной сфере такие информационно-исследовательские системы (ИИС) опираются на базовые научные онтологии и паттерны онтологического проектирования (Ontology Design Patterns или ODP) [1-3], которые позволяют сэкономить усилия разработчиков. Этот подход закладывает основу создания экспертно-ориентированной [4] технологии разработки ИИС, когда, с одной стороны, эксперту предоставляются необходимые инструменты для интегрирования экспертных знаний в базу знаний системы, а с другой стороны, появляется возможность создавать собственные ИИС без участия программистов.

В статье предлагается подход к созданию конструктора ИИС, обеспечивающего разработку онтологий и интеграцию разнородных ресурсов в единое информационное пространство, а также настройку визуализации элементов онтологии и базы знаний системы в виде многостраничного сайта.

Дальнейшая структура изложения включает следующие вопросы. В разделе 1 приведен обзор предшествующих работ. В разделе 2 приведена модель конструктора ИИС, описывающая все необходимые компоненты разрабатываемой технологии. В разделе 3 приведены примеры применения конструктора для создания простых ИИС, демонстрирующие основные возможности созданной технологии. В разделе 4 обсуждаются некоторые ключевые возможности и проблемы.

1. История вопроса. Существует много исследований, направленных на создание технологии разработки ИИС на базе онтологий. В работе [5] отмечается, что в настоящее время, несмотря на то, что онтологии используются в большом количестве ИИС, разработчики новых систем сталкиваются с рядом существенных проблем, поскольку эти две области – область разработки онтологий и область разработки программного обеспечения, развивались, как правило, раздельно. Системы, предназначенные для проектирования и разработки программных систем под управлением онтологии, относятся к области *Ontology-Based Software Engineering* [6, 7]. Часто такие системы предоставляют функциональность, схожую с Protégé, поскольку снабжены инструментами разработки и редактирования онтологий.

В рамках данной работы мы развиваем подход, представленный в [8, 9]. Идея этого подхода заключается в организации доступа к систематизированным знаниям и информационным ресурсам конкретной предметной области путем использования онтологии в качестве его концептуальной основы, а также использование различных методик для автоматизации разработки онтологий. В соответствии с предлагаемым подходом авторами были разработаны редактор данных, модули поиска и навигации, автоматического индексирования и классификации данных, а также создан ряд ИИС по различным отраслям науки [10, 11].

При разработке ИИС для предметной области «Фольклор народов Сибири» дополнительно к задачам, связанным с моделированием предметной области, были поставлены задачи: а) семантической разметки текстов, б) поддержки параллельных текстов на разных языках, в) поддержки мультимедийных ресурсов (аудио, видео, графические изображения). В работе [12] было приведено описание подхода к разработке ИИС, в рамках которого система должна обеспечивать создание и поддержку репозитория информационных ресурсов, аннотирование ресурсов на основе предметных онтологий, а также создание и поддержку предметных онтологий на уровне простых операций редактирования и обеспечение доступа к ресурсам и данным.

В данном походе, роли онтологий разделены:

1. Онтология ресурсов – формальное описание коллекции информационных ресурсов, представленных в ИИС; является в некотором смысле метаонтологией представления и визуализации, поскольку формирует основу модели данных для пользовательского интерфейса, навигации и поиска.
2. Предметная онтология – формальное описание сущностей предметной области, которое является основой для аннотирования содержащихся в ИИС ресурсов.

В основу онтологии ресурсов легла формальная модель, созданная для поддержки интеграции, стандартизации и обмена информацией в области культурного наследия – CIDOC-CRM [13]. С помощью «Событий», являющихся основными концептами данной онтологии, выражаются процессы сбора и подготовки фольклорных произведений. Также добавлены некоторые свойства, отсутствующие в оригинальной онтологии CIDOC-CRM, для более естественного представления информации, где некоторые из них импортированы из онтологий FOAF, DBpedia и GeoNames (foaf:firstName, dbo:title, geo:lat, geo:long и т.д.). Несмотря на гибкость используемой модели, возникает ряд проблем при непосредственной реализации функций создания, удаления и редактирования узлов, связанных с загружаемыми

ресурсами. Из-за жесткого ориентирования модели на понятие «Событие», внесение любых изменений в метаинформацию файлов системы становится сложным процессом из-за многочисленных повторяющихся связей.

Модуль предметно-ориентированной разметки параллельного текста [14] обеспечивает как синхронизацию переводного текста с оригиналом (построчно), так и аннотирование специалистом сущностей ПрО в тексте. Модель аннотации текста включает множество «вхождений» элементов онтологии в текст или множество фрагментов, размеченных элементами онтологии – экземплярами классов и отношений. В то же время модель не поддерживала разметку отдельных фрагментов текста – части строк или целых абзацев, а только строк, что было связано с особенностями представления фольклорного текста.

Дальнейшее развитие ИИС заключалось в использовании графовых баз данных для размещения онтологий предметных областей и онтологии ресурсов [15]. Графовые базы данных предоставляют более обобщенную структуру, основанную на модели LPG – Labeled Property Graphs (графы меток и атрибутов) [16]. Элементы онтологий обладают уникальными идентификаторами (URI), позволяющими однозначно идентифицировать их принадлежность. В графовом представлении (LPG) эту функцию берут на себя метки.

Пространство имен соответствующей онтологии также включается в описание вершин графа, что позволяет хранить множество онтологий в одном графе. При этом одни и те же вершины могут одновременно быть частью более чем одной онтологии. На основе LPG была разработана модель трансляции онтологии в графовую базу данных, а также система иерархии онтологий с использованием специальных меток узлов «Secondary» и «Exclude» (онтология, которая не наследуется от других, называется корневой, а метки присваиваются расширенным и ограниченным онтологиям соответственно). Несмотря на предоставляемую гибкость в наследовании онтологий, данный метод негативно влиял на время выполнения запросов к базе данных, сильно нагружая систему, а также затруднял разделение одной онтологии от другой без использования специальных инструментов визуализации данных.

На основе разработанных инструментов была создана электронная коллекция фольклорных материалов [17], собранных в Новосибирской области и опубликованных в период с 1969 по 2021 годов. Данная ИИС поддерживает следующие функции:

- хранение и редактирование репозитория информационных ресурсов;
- аннотирование ресурсов на основе предметных онтологий;
- создание и поддержка предметных онтологий на уровне простых операций;
- обеспечение доступа к ресурсам и данным;
- навигация.

Использование конкретной ПрО при проектировании и создании ИИС предусматривает жесткую структуру модели онтологии ресурсов, а также модулей по интеракции конечных пользователей с системой. Следующим шагом в работе с ИИС становится отход от конкретных предметных областей в пользу адаптивных систем по редактированию онтологии ресурсов, а также детальной настройки элементов системы не ее разработчиками, а ее экспертами. Такой подход рассматривается в работе [18]. Авторы предлагают технологию построения ИИС, ориентированную на экспертов. В данном решении база знаний предоставляет знания о ПрО, в рамках которой будет функционировать система: основные сущности (понятиях и объектах) ПрО, а также связи между ними. К последним относятся знания о связывающих понятия отношениях, а также более сложные зависимости между понятиями ПрО. Следующий шаг в развитии адаптивных систем был предложен в работе [19], которая описывает метод создания автоматически генерируемых адаптивных пользовательских интерфейсов редакторов баз знаний, построенных на основе

онтологического подхода. Эта модель подразумевает использование заранее заданных абстрактных интерфейсных элементов, описывающих конкретные блоки ввода и вывода данных из базы знаний: списки выбора, логические переменные, текстовые поля, кнопки и т.д. Несмотря на гибкость данного решения, настройка лишь одной формы в рамках разработанной системы требует большого количества времени и детальной настройки, что может привести к ошибкам типизации и вводу некорректных данных в базу знаний.

В соответствии с требованиями к конструктору ИИС был составлен список основных задач, содержащий такие пункты как создание модели платформы-конструктора ИСС, обновление общей архитектуры и разработку новых инструментов работы с модулями отображения ресурсов. Разрабатываемая система должна реализовывать следующие функциональные требования:

- создание и редактирование ИИС;
- создание и редактирование web-страниц ИИС;
- создание, редактирование и наследование онтологий;
- загрузка, редактирование и отображение мультимедиа ресурсов;
- семантическая разметка загруженных текстов (в том числе параллельных).

2. Конструктор информационно-исследовательских сред. В основе конструктора информационных сред лежит базовая (основная) онтология ресурсов. Задача этой онтологии заключается в представлении корневых понятий системы в виде ее элементов – связи между файлами, их типами и узлами, к которым прикрепляются данные, а также список поддерживаемых форматов.

Конструктор представляет множество загруженных или созданных в нем онтологий предметных областей в виде иерархии, с которой можно взаимодействовать непосредственно через блок редактора. Блок редактора паттернов обеспечивает редактирование, создание и удаление паттернов соответственно.

Блок семантической разметки работает с реляционной базой данных для хранения связей между конкретными фрагментами текста и сущностями или экземплярами онтологии предметных областей.

Модель конструктора можно формально описать в следующем виде:

$$IE = \langle BORes, IBOS, PO, OntoEditor, PagesEditor, TextAnnotator \rangle \quad (1)$$

где *BORes* – базовая онтология ресурсов, *IBOS* – иерархия базовых онтологий ПрО, *PO* – множество онтологических паттернов, *OntoEditor* – редактор онтологий и паттернов, *PagesEditor* – редактор страниц системы (информационной среды), *TextAnnotator* – редактор семантической разметки текстов.

С помощью конструктора создается проект, который представляет собой набор основных визуальных элементов – редакторы онтологий и галерею ресурсов, а также множество настраиваемых страниц, созданных пользователем.

В основе проекта лежит выбранная при его создании основная (корневая) онтология ресурсов, а также множество онтологий предметных областей для разметки текстов и проведения исследовательских работ в предусмотренной области.

Формально, проект описывается системой вида:

$$IS = \langle ORes, OS, MPages, MBs, FPB, FB \rangle \quad (2)$$

где *ORes* – онтология ресурсов, *OS* – множество онтологий предметных областей,

MPages – множество настраиваемых страницы,

MBs – множество визуальных элементов,

FPB – функция, которая сопоставляет и задает правила визуализации PV элементов онтологий:

$$FPB = MBs \rightarrow (ORes \cup OS) \times PV \quad (3)$$

FB – функция, которая сопоставляет элемент и страницу IS и задает правило визуализации элемента на странице:

$$FB = MBs \rightarrow MPages \times PV \quad (4)$$

2.1. Иерархия онтологий. Иерархия онтологий в ИИС позволяет расширять и ограничивать уже существующие онтологии без необходимости полного дублирования их содержания. С помощью данной функции пользователь может добавлять новые элементы в любую онтологию, представленную в библиотеке конструктора, путем наследования от данной онтологии.

В первой модели для выделения онтологической иерархии используются метки «Secondary» и «Exclude» для отображения расширенной и ограниченной онтологий. Эти метки присваивались всем классам онтологии. В результате апробации системы на нескольких предметных областях был сделан вывод об ее нагруженности и плохой производительности.

Во второй модели механизм иерархии основан на единственном поле в узле онтологии под названием «signature». Это поле хранит в себе массив строк, каждая из которых является URI-меткой определенной онтологии. Пример представлен на рисунке 1.

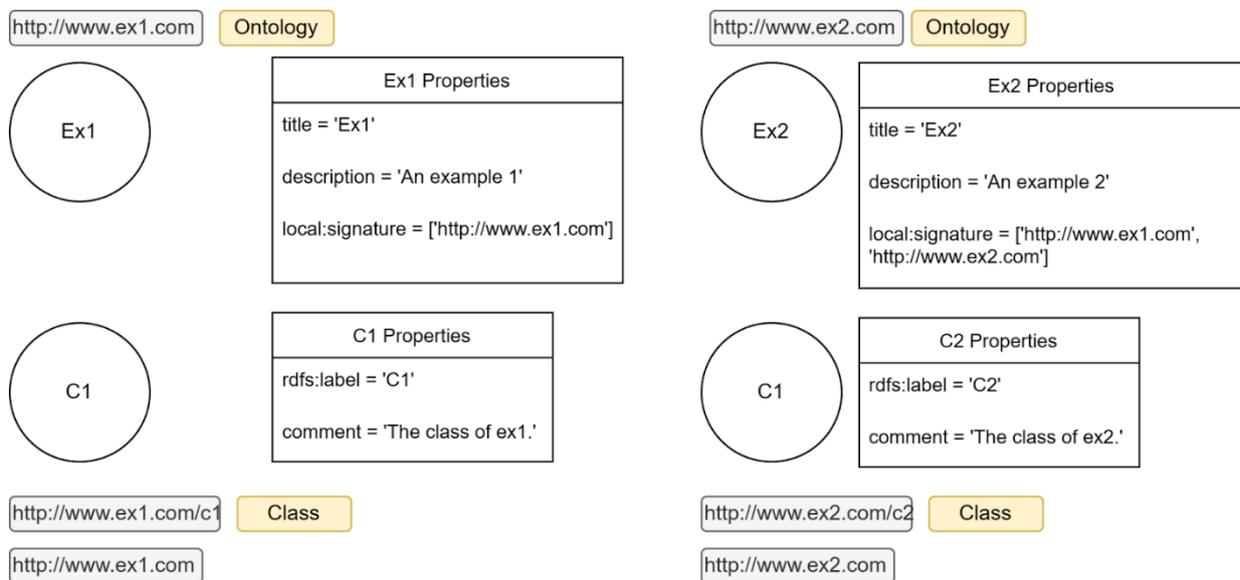


Рис. 1. Пример сопоставляется онтологий узлам графа

В примере (рис. 1) приведена корневая онтология «Ex1» и ее наследник «Ex2».

При запросе элементов определенной онтологии, вместо выполнения запроса по ее URI, выполняется запрос по ее полю «signature». Так как в корневой онтологии данное поле равно ее URI, то при ее запросе будут возвращены только ее элементы.

В случае онтологии «Ex2» ее поле «signature» содержит не только собственный URI, а также URI онтологии «Ex1», поэтому, при запросе элементов онтологии «Ex2», пользователю будут возвращены элементы сразу из двух онтологий.

В результате апробации новой модели было отмечено, что по сравнению с предыдущей версией среднее время запроса к графовой базе данных было ускорено примерно в 1,5 раза.

2.2. Онтология ресурсов. Корневая онтология ресурсов представляет собой мета-онтологию для поддержки интеграции проектов системы и их стандартизации для обмена данными. На основе этой онтологии конечный пользователь имеет возможность построить свою онтологию ресурсов путем наследования компонентов из корневой, тем самым предлагая свою спецификацию мета-онтологии для конкретной предметной области.

Скелет корневой онтологии был разработан на основе CIDOC-CRM [13]. Концептуальная эталонная модель CIDOC (CRM) представляет собой теоретический и практический инструмент для интеграции информации в области культурного наследия. Она позволяет исследователям, администраторам и общественности исследовать сложные вопросы, касающиеся нашего прошлого, в разнообразных и рассредоточенных наборах данных. CIDOC CRM достигает этого, предоставляя определения и формальную структуру для описания неявных и явных понятий и отношений, используемых в документации по культурному наследию и представляющих общий интерес для запроса и исследования таких данных. Такие модели также известны как формальные онтологии. Эти формальные описания позволяют интегрировать данные из нескольких источников независимо от программного обеспечения и схемы.

CIDOC CRM является результатом более чем 20-летней работы по разработке и сопровождению, первоначально проводившейся Рабочей группой по стандартам документации CIDOC, а в настоящее время – группой CIDOC CRM SIG, которые являются рабочими группами CIDOC. С 2006 года он признан официальным стандартом ISO [20].

Графовое представление базовой онтологии ресурсов представлено на рисунке 2.

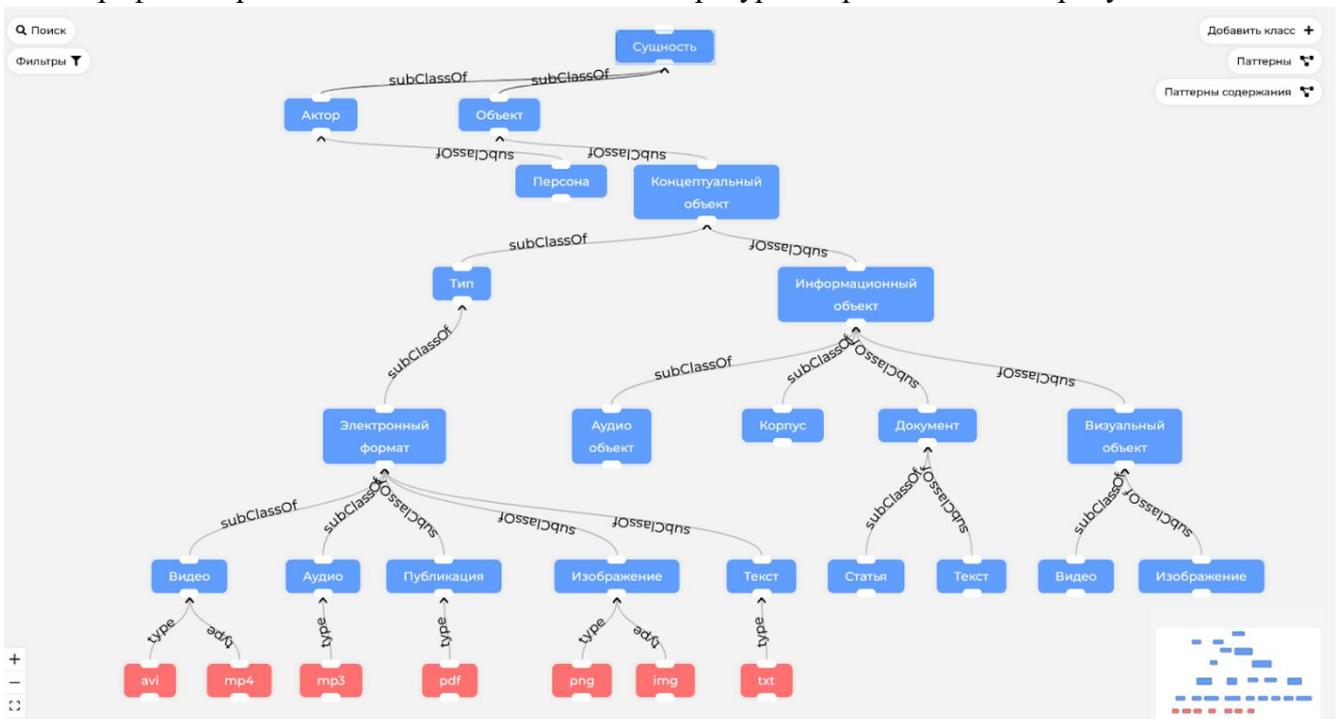


Рис. 2. Базовая онтология ресурсов, основанная на CIDOC-CRM

Для обеспечения поддержки множества предметных областей в онтологии были проведены следующие модификации:

- базовая онтология не содержит концепты для описания временных явлений (Temporal Entities – Event): «Постоянный элемент» убран, Актор и Предмет расположились выше по иерархии;
- класс «Объект» содержит в себе только Концептуальные и Материальные объекты;
- класс Appellation убран в пользу связи Информационный объект -> определяет -> Сущность и атрибута «Файл» класса «Информационный объект»;
- такие классы, как Жанр, Валюта, Меры измерения и т.д. были исключены из базовой онтологии по причине их специфичности.

Базовая онтология ресурсов (метаонтология) представлена на рисунке 2. Свойства подклассов класса «Информационный объект» представлены на рисунке 3.

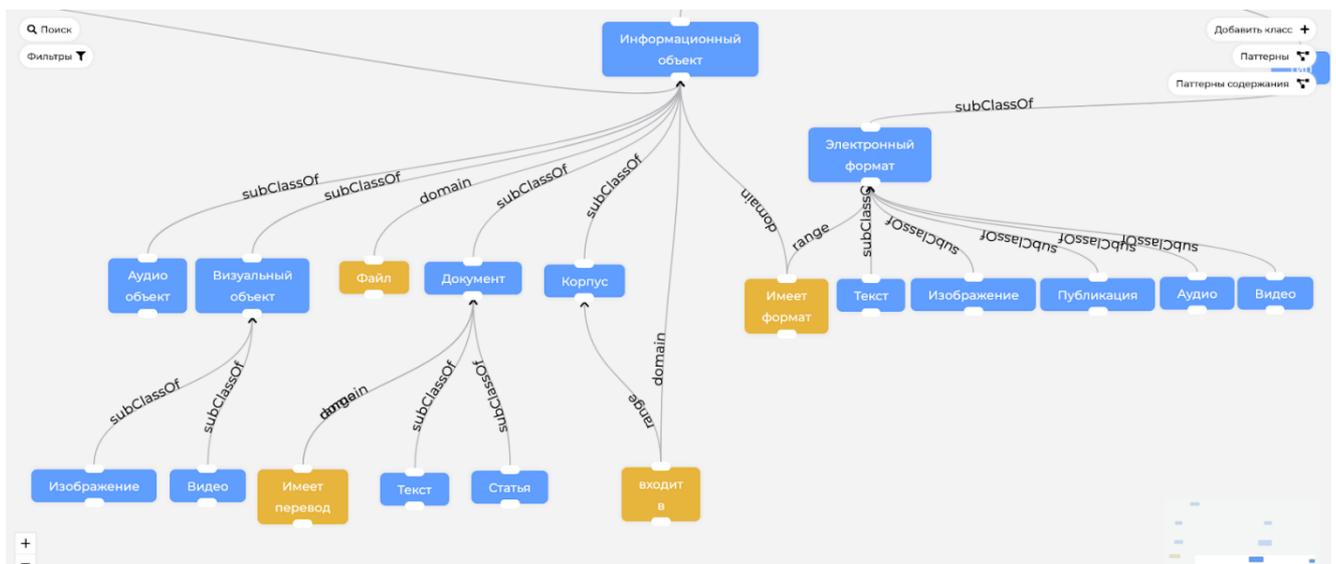


Рис. 3. Свойства подклассов класса «Информационный объект»

2.3. Семантическое аннотирование ресурсов. Для внедрения системы семантической разметки, а также для разработки системы визуализации мультимедиа элементов онтологии ресурсов в основной онтологии ресурсов была создана сущность «Корпус».

Термин «Корпус текстов» используется в мировой лингвистике на протяжении многих лет. В широком смысле под этим термином понимают группу текстов, собранных и организованных в соответствии с какой-либо целью или замыслом. В частном случае под замыслом понимают использование корпуса в качестве базы для исследования того или иного языка: статистического анализа, проверки статистических гипотез и подтверждения лингвистических правил

Сущность «Корпус» является наследником класса «Информационный объект», так же, как и сущность «Документ». Сущность «Документ» включает подкласс «Текст», который хранит в себе файлы формата txt. Сущность «Документ» может определять любой класс мультимедийных объектов, и связь «Входит в» привязывает данные объекты непосредственно к корпусу.

Помимо сущности «Текст», были созданы сущности, корреспондирующие к каждому типу мультимедиа объектов: «Статья», «Видео», «Изображение», «Аудио объект».

Структура созданных классов относительно корневой онтологии представлена на рисунке 3.

2.4. Конструктор страниц. Конструктор страниц представляет собой набор инструментов, предназначенный для создания и редактирования специальных страниц ИИС. Его основная функция заключается в предоставлении конечным пользователям возможности полноценной кастомизации их проекта.

Под кастомизацией в данном случае понимается процесс изменения внешнего вида и функциональности веб-страниц, с целью сделать их более индивидуальными и удобными для пользователей.

Конструктор разделен на следующие структурные элементы (рисунок 4):

- проект,
- страница,
- блок.

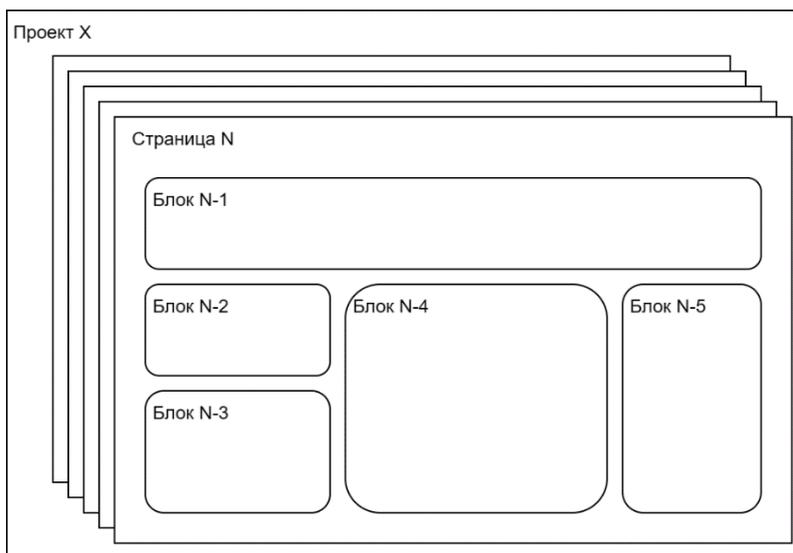


Рис. 4. Структурная модель ИИС

«Проект» представляет собой объект, хранящий ключевую информацию об информационном портале. Он включает в себя список онтологий предметных областей, название и описание проекта, а также ссылку на унаследованную онтологию ресурсов для хранения и представления мультимедиа материалов.

«Страница» – часть «проекта», состоящая из «блоков».

«Блок» – объект, описывающий определенный функциональный элемент. Описание содержит координаты элемента на странице, а также информацию, специфичную для соответствующего типа блока.

Типы блоков: список, мозаика, изображение, слайд-шоу, текст.

Конфигурация блока представлена в листинге 1.

Листинг 1. Конфигурация блока

```

TPageBlock = {
  page_id: number,
  id?: number,
  block_type: 'text' | 'card_list' | 'bullet_list' | 'media' | 'media_slide_show',
  list_uri?: '',
  x: number,
  y: number,
  h: number,
  w: number,
  data?: {
    // picture
    picture_data?: string,
    // text
    text?: string,
    font_size?: number,
    font_weight?: boolean,
    ontology_uri?: string,
    // card_list - bullet list
    class_uri?: string,
    class_attributes?: string[],
  }
}
    
```

В данном листинге используются следующие обозначения: x, y, h, w – координаты блока на странице и его ширина и высота; picture_data – ссылка на изображение блока «медиа»; text, font_size, font_weight – текст, размер шрифта и жирность шрифта блока «text»; ontology_uri, class_uri, class_attributes – ссылка на онтологию; ссылка на класс и список отображаемых атрибутов для блока типа «список» или «мозаика».

С помощью данного инструмента можно создавать такие страницы, как списки услуг или сотрудников, главные страницы или страницы контактов организации, а также галереи для просмотра мультимедиа файлов проекта.

2.5. Авторизация. Авторизация пользователей системы реализована с помощью инструмента VK ID [21]. VK ID представляет собой платформу, обеспечивающую быструю интеграцию форм входа в произвольный разрабатываемый сервис. Процесс авторизации заключается в получении уникального идентификатора пользователя от платформы VK, совместно с его инициалами. Далее полученные данные записываются в объект *Account*, унаследованный от модели *Пользователь* фреймворка *Django*, для которого автоматически генерируется почтовый ящик и пароль. Генерация пароля не влияет на процесс авторизации, так как его ключом является токен, полученный со стороны VK ID.

3. Экспериментальное исследование. Для апробации новой информационной структуры был создан тестовый корпус «Поэты» и наполнен примерами статей, видеоматериалами и стихотворениями для разметки.

3.1. Наследование от корневой онтологии ресурсов. Наследование онтологии ресурсов «Поэты (рес.)» осуществляется от корневой онтологии ресурсов (метаонтологии) (рисунок 5).

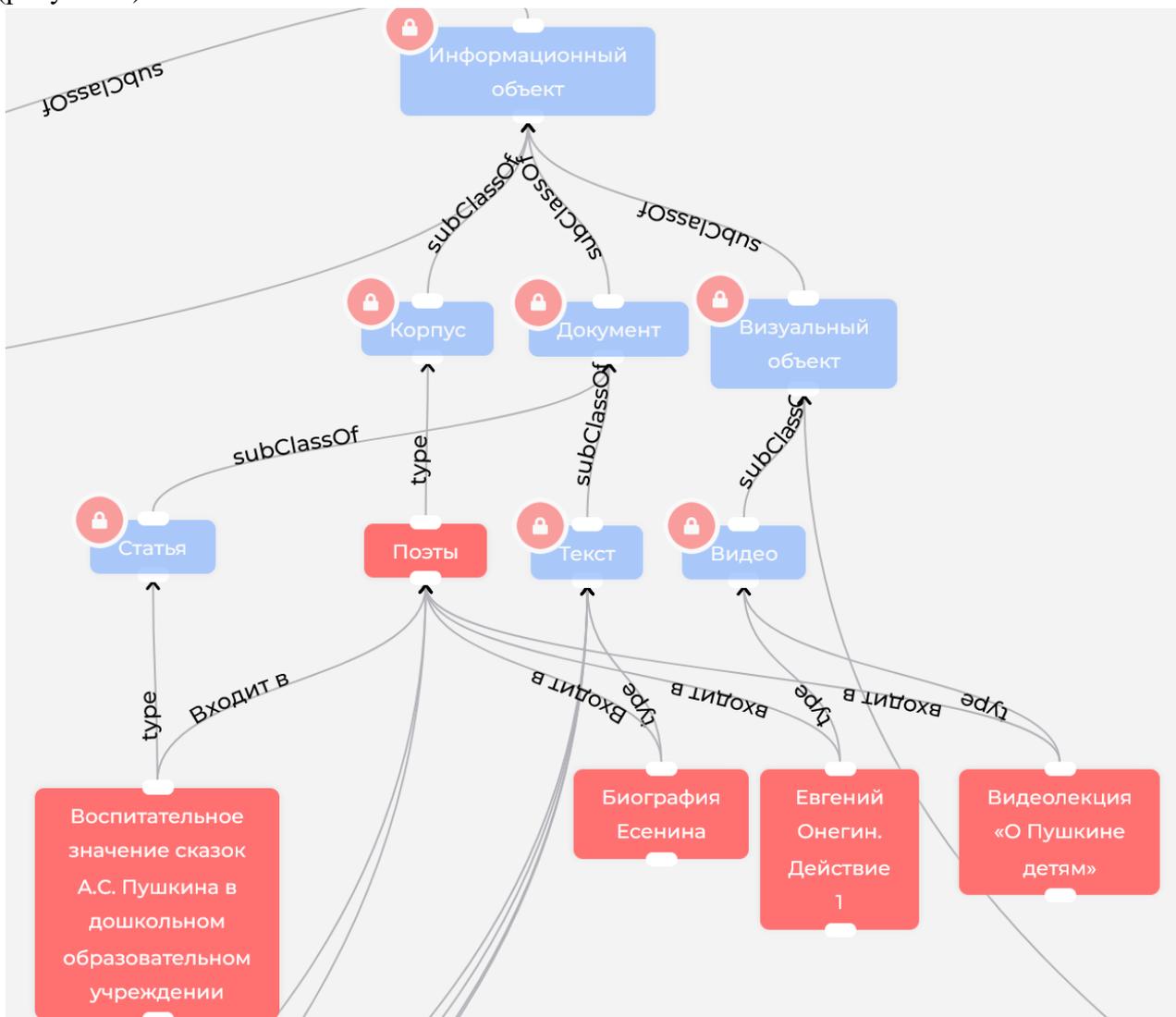


Рис. 5. Визуализация механизма наследования онтологии

На рисунке 5 представлено, как в графическом интерфейсе текущая онтология (Поэты) отделяется от конечной посредством меток замка и регулировкой цветовой палитры.

3.2. Создание ИИС и настраиваемых страниц. Процесс создания нового ресурса заключается в выборе пользователем необходимых онтологий ПрО, а также онтологии ресурсов (метаонтологии), на основе которых будет построена структура системы. Форма создания нового информационного ресурса представлена на рисунке 6.

Проект

Название: Поэты

Онтологии ресурсов: Онтология ресурсов Поэты(рес.)

Онтологии: Базовая Поэты Ассистенты Компьютерная лингвистика (база) Компьютерная лингвистика

Рис. 6. Форма создания нового информационного ресурса

Для визуализации ресурсов корпуса была создана и добавлена настраиваемая страница «Поэты». На рисунке 7 представлена страница настройки отображения каталога загруженных ресурсов.

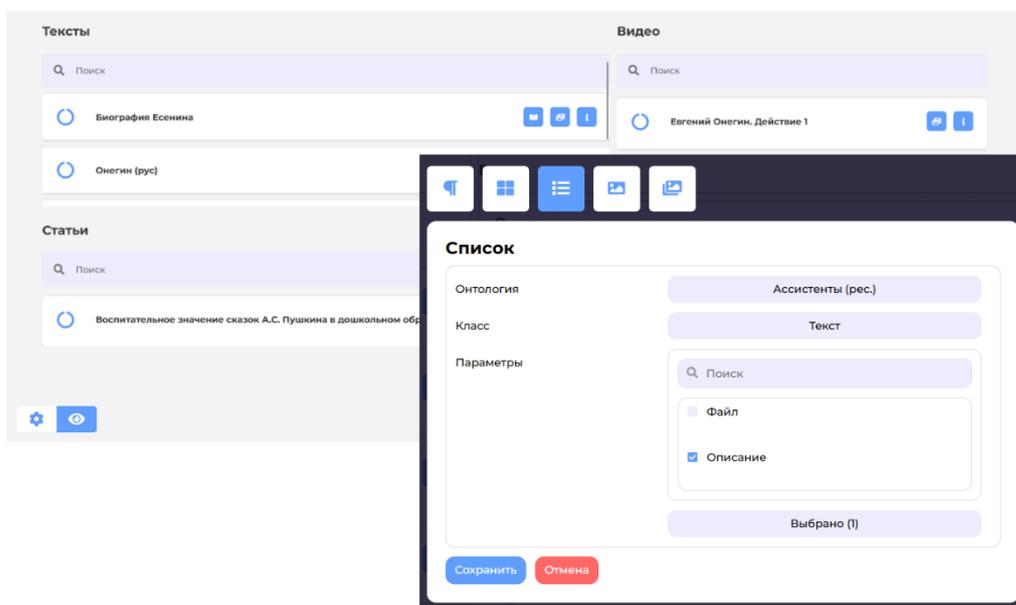


Рис. 7. Каталог ресурсов созданного проекта

На данной странице пользователь с помощью шести функциональных блоков (трех блоков типа «Текст» и трех блоков типа «Список») обеспечил отображение списков текстов, видеоматериалов и статей. При настройке отображения блока «Текст» с помощью формы, представленной на рисунке 7, были выбраны параметры визуализации на основе онтологических свойств.

3.3. Семантическая разметка текстов. За основу модели инструмента для семантической разметки было взято решение, разработанное в 2021 г. при создании системы «Веб-платформа для фольклорных исследований на основе онтологии предметных областей» [10].

Для поддержки аннотирования и отображения параллельных текстов в класс «Текст» был добавлен атрибут «Имеет перевод», связывающий тексты на разных языках.

В функциональные блоки были добавлены специальные ссылки для перехода на страницу аннотирования текстов. Эти ссылки отображаются только в файлах формата txt и определяют, какой из двух связанных текстов является основным. Для примера, если конечный пользователь системы откроет текст «Онегин (англ.)», то данный текст будет использован для аннотирования, имея текст «Онегин (рус.)» в качестве параллельного. Также предусмотрен обратный вызов.

На рисунке 8 представлены фрагменты размеченного текста с подсветкой различных связей.

0	Всего , что знал еще Снегин Евгений ,	0	Всего , что знал еще Снегин Евгений ,	0	Всего , что знал еще Евгений ,
1	Пересказать мне недосуг;	1	Пересказать мне недосуг;	1	Пересказать мне недосуг;
2	Но в чем он истинный рождение был гений ,	2	Но в чем он истинный имеет характеристику был гений ,	2	Но в чем он истинный был гений ,
3	Что знал он тверже всех наук ,	3	Что знал он тверже всех наук ,	3	Что знал он тверже всех наук ,
4	Что было для него измлада	4	Что было для него измлада	4	Что было для него измлада
5	И труд и мука и отрада ,	5	И труд и мука и отрада ,	5	И труд и мука и отрада ,
6	Что занимало целый день	6	Что занимало целый день	6	Что занимало целый день
7	Его тоскующую лень , -	7	Его тоскующую лень , -	7	Его тоскующую лень , -
8	Была наука страсти нежной ,	8	Была наука страсти нежной ,	8	Была наука страсти нежной ,
9	Которую воспел Назон ,	9	Которую воспел Назон ,	9	Которую воспел Назон ,
10	За что страдалец умер в кончил он	10	За что страдалец кончил он	10	За что страдалец умер в кончил он
11	Свой век блестящий и мятежный	11	Свой век блестящий и мятежный	11	Свой век блестящий и мятежный
12	В Молдавия , в глуши степей ,	12	В Молдавия , в глуши степей ,	12	В Молдавия , в глуши степей ,

Рис. 8. Фрагменты размеченного текста с визуализацией связей

Также была реализована функция по стандартному аннотированию текстов, не имеющих перевода.

На рисунке 9 изображен процесс привязки сущности или экземпляра к фрагменту текста. Процесс привязки включает в себя выделение фрагмента размечаемого текста и активации формы выбора элемента онтологии, который необходимо связать с этим фрагментом. Форма разделена на несколько типов элементов онтологии: классы, объекты и именные связи. Блоки выбора классов, объектов и связей содержат функции полнотекстового поиска.

3.4. Визуализация мультимедиа ресурсов. Для просмотра мультимедиа был создан новый функциональный компонент, подстраиваемый под отображаемый им тип файла. Под самим файлом находится его описание в соответствии с информацией в его узле в онтологии ресурсов. На рисунке 10 представлены примеры визуализации статьи, видео и текста соответственно.

3.5. Универсальность конструктора. Для демонстрации универсальности разработанной системы был создан дополнительный проект, предметная область которого содержит информацию о сотрудниках, их проектах и публикациях лаборатории Искусственного интеллекта Института систем информатики СО РАН.

Для загрузки данных были созданы 3 новых класса в унаследованной онтологии ресурсов: *Сотрудник*, *Публикация*, *Проект*. Структура классов представлена на рисунке 11.

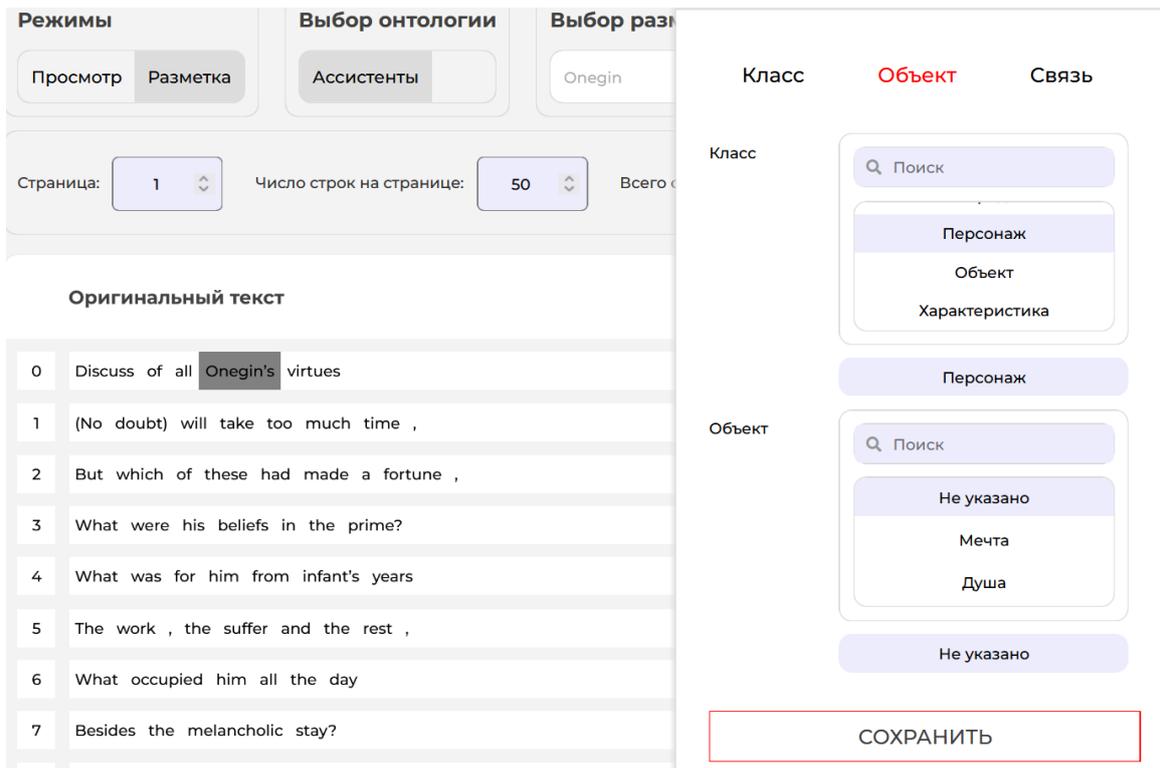


Рис. 9. Фрагмент текста и форма выбора типа элемента онтологии для связи

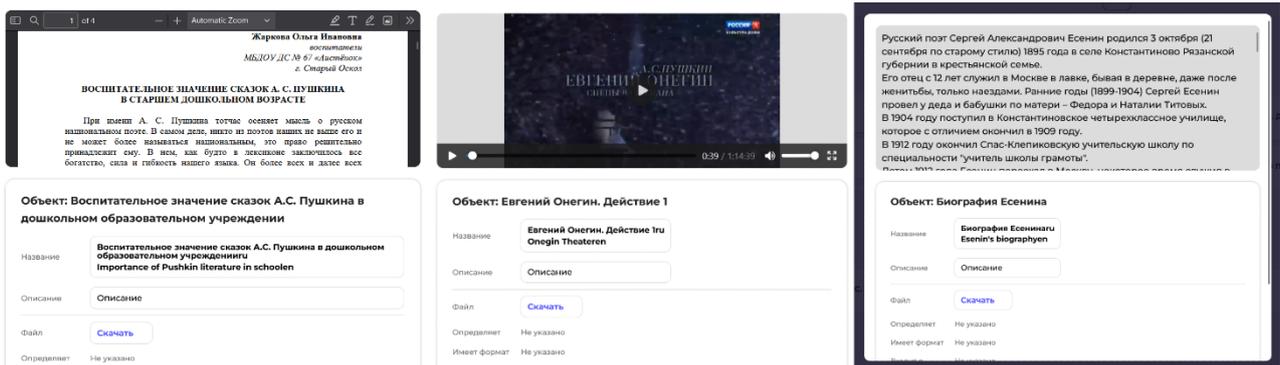


Рис. 10. Пример отображения содержания статьи, видеоматериала и текста

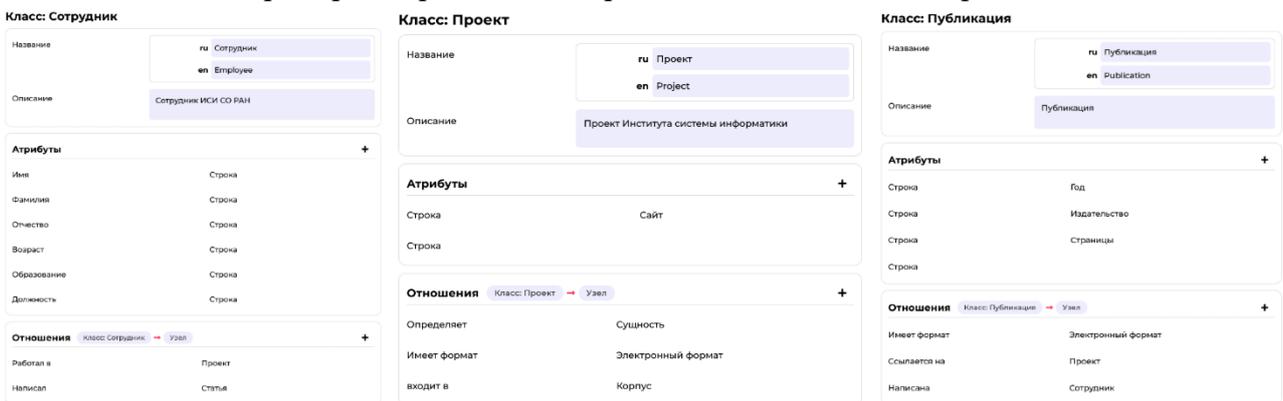


Рис. 11. Интерфейс редактирования объектов классов онтологии

В систему была загружена информация о 10 сотрудниках, 10 проектах и 32 публикациях. Для каждого класса с помощью конструктора страниц были созданы персонализированные страницы. Страница «Сотрудники лаборатории» представлена на рисунке 12.

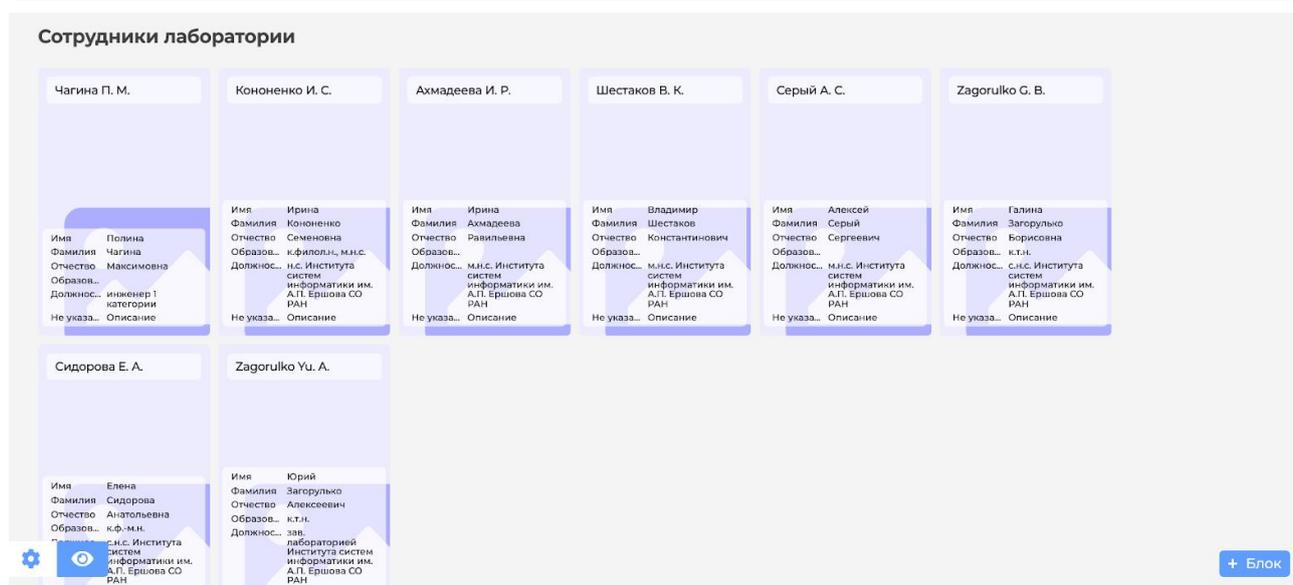


Рис. 12. Страница «Сотрудники»

4. Обсуждение результатов. В представленной статье авторы пытаются решить ряд сложных методологических проблем.

Первая проблема – обеспечение end-to-end конструирования систем, – широко известна и имеет множество решений, таких, как WordPress или Squarespace. Приведенные технологии, несомненно, предоставляют пользователям широкий набор инструментов для создания и редактирования web-ресурсов, не требуя от них глубоких знаний в программировании. Несмотря на это, данные решения используются и предназначены для малых и одностраничных приложений, ориентированных на бизнес-продукты, такие, как продажа, реклама малых бизнесов или персональные блог-страницы. Авторами предлагается решение именно для класса ИИС, основанных на онтологиях, для проведения исследований в конкретных предметных областях. Ключевыми компонентами такого решения являются механизмы повторного использования базовых онтологий и применения иерархий на онтологиях, а также принципиальная возможность создания и внедрения библиотеки паттернов онтологического проектирования.

Вторая проблема – семантическая связь ресурсов и предметных понятий. Типовым решением является описание мета-параметров ресурса: *автор*, *дата создания*, *тематика* и т.п., но не анализ содержания. Многие ресурсы обладают собственным богатым содержанием: тексты, видео, аудио. Авторы разработали систему семантического аннотирования для текстов, связывающую фрагменты текста с соответствующими сущностями и экземплярами прикрепленных онтологий.

И, наконец, проблема конструирования многостраничного ресурса с индивидуальными настройками визуализации. Ранее в [14, 15] рассматривались только локальные настройки визуализации сущностей ПрО. В данной работе предложено более универсальное решение, позволяющее в рамках структурной модели сайта связать страницу с онтологическим классом(ами) и выбрать способы визуализации онтологических элементов.

Недостатки предложенного решения заключаются в громоздкости представления данных для больших онтологий. Эта проблема распространяется на все решения, основанные на визуализации больших графов. Решение может быть связано с реализацией «умной» свертки подграфов, а также с улучшением алгоритмов изначального позиционирования узлов. В итоге, необходимо достичь компромисса между детальной визуализацией всех элементов

онтологии на графе и имплементации плавного, отзывчивого и дружелюбного интерфейса для конечных пользователей системы.

Заключение. Целью работы является создание конструктора ИИС, обеспечивающего разработку онтологий и интеграцию разнородных ресурсов в единое информационное пространство, а также настройку визуализации элементов онтологии и базы знаний системы в виде многостраничного сайта. В статье приведены описания подхода, модели ИИС, модели иерархии онтологий, а также примеры апробации разработанной системы для двух предметных областей с примерами разметки текстов, визуализации онтологий в виде интерактивных графов, а также примеров персонализировано-настроенных блоков интерфейса с помощью созданного конструктора страниц.

В будущем планируются оптимизация и улучшение алгоритмов визуализации графов онтологий, разработка библиотеки онтологических паттернов и использование методов RAG (Retrieval Augmented Generation) для реализации поиска с помощью языковых моделей.

Список источников

1. Gangemi A., Presutti V. *Ontology design patterns*. Handbook on ontologies, Springer, 2009, pp. 221-243, DOI: 10.1007/978-3-540-92673-3_10.
2. Загорулько Ю.А. Применение паттернов онтологического проектирования при разработке онтологий научных предметных областей / Ю.А. Загорулько, О.И. Боровикова, Г.Б. Загорулько // Аналитика и управление данными в областях с интенсивным использованием данных: сборник научных трудов XIX Международной конференции DAMDID/RCDL'2017. – Москва: ФИЦ ИУ РАН, 2017. – С. 332-340.
3. Ломов П.А. Применение паттернов онтологического проектирования для создания и использования онтологий в рамках интегрированного пространства знаний / П.А. Ломов // Онтология проектирования, 2015. – Т. 5, – № 2(16). – С. 233-245. – DOI:10.18287/2223-9537-2015-5-2-233-245.
4. Дрождин В.В. Экспертно-ориентированный подход к разработке прикладного программного обеспечения / В.В. Дрождин, М.В. Жуков // Проблемы информатики в образовании, управлении, экономике и технике: Сб. статей IX Междунар. научно-техн. конф. – Пенза: ПДЗ, 2009. – С. 131-135.
5. Alobaid A., Garijo D., Poveda-Villalón M., et al. Automating ontology engineering support activities with OnToology. *Journal of Web semantics*, 2019, no. 57, pp. 100472, DOI:10.1016/j.websem.2018.09.003.
6. Isotani S., Bittencourt I.I., Barbosa E.F., et al. Ontology driven software engineering: a review of challenges and opportunities. *IEEE Latin America transactions*, 2015, no. 13(3), pp. 863-869, DOI: 10.1109/TLA.2015.7069116.
7. Хорошевский В.Ф. Проектирование систем программного обеспечения под управлением онтологий: модели, методы, реализации / В.Ф. Хорошевский // Онтология проектирования, 2019. – Т. 9. – № 4(34). – С. 429-448.
8. Zagorulko Y.A., Zagorulko G.B. Experience of using content patterns in the development of Ontologies of scientific subject areas. *Pattern recognition and image analysis*, 2023, vol. 33, iss. 3, pp. 560-567, DOI:10.1134/S1054661823030513.
9. Загорулько Ю.А. Программный комплекс для разработки интеллектуальных систем поддержки принятия решений в слабоформализованных областях / Ю.А. Загорулько, Г.Б. Загорулько, А.С. Серый и др. // Наука и технологии Сибири. Сибирская индустрия информационных систем, 2023. – № 10. – С. 34-38.
10. Sapetina A., Kulikov I., Zagorulko G., et al. Constructing an expert system for solving astrophysical problems based on the ontological approach. *Parallel computational technologies, communications in computer and information science*, Springer, 2022. vol. 1618, pp. 30-42, DOI:10.1007/978-3-031-11623-0_3.
11. Zagorulko Y., Borovikova O., Zagorulko G. Knowledge portal on computational linguistics: content-based multilingual access to linguistic information resources. *Selected topics in Applied computer science. Proceedings of the 10th WSEAS International Conference on Applied Computer Science (ACS'10)*. Hamido Fujita, Jun Sasaki (Eds.), WSEAS Press, 2010, pp. 255-262.
12. Серый А.С. Подход к построению информационной исследовательской среды для анализа научных материалов и создания аннотированных корпусов / А.С. Серый, А.А. Гриневиц, В.А. Лисин // Информационные и математические технологии в науке и управлении, 2021. – № 4(24). – С. 68-76.
13. CIDOC Conceptual reference model, available at: <http://www.cidoc-crm.org/> (accessed: 12/01/2021).
14. Лисин В.А. Веб-платформа для фольклорных исследований на основе онтологии предметных областей / В.А. Лисин, Е.А. Сидорова // Вестник Новосибирского государственного университета. Серия: Информационные технологии, 2021. – Т. 19. – № 2. – С. 53-64.

15. Лисин В.А. Модель представления онтологии предметных областей на основе графовых баз данных / В.А. Лисин, А.С. Серый, Е.А. Сидорова // Вестник Новосибирского государственного университета. Серия: Информационные технологии, 2022. – Т. 20. – № 4. – С. 24-38. – DOI: 10.25205/1818-7900-2022-20-4-24-38.
16. Jesús Barrasa RDF Triple stores vs. Labeled property graphs: what's the difference? available at: <https://neo4j.com/blog/rdf-triple-store-vs-labeled-propertygraph-difference> (accessed: 01/13/2023).
17. Гриневиц А. Анализ культурных универсалий фольклора народов Сибири и Дальнего Востока / А. Гриневиц, А. Серый // Материалы XXIII Международной конференции по аналитике и управлению данными в информационно емких областях (DAMDID/RCDL 2021). CEUR, 2021. Vol. 3036. P. 387–401.
18. Загоруйко Ю.А. Семантическая технология разработки интеллектуальных систем, ориентированная на экспертов предметной области / Ю.А. Загоруйко // Онтология проектирования, vol. 5, no. 1 (15), 2015. – С. 30-46.
19. Грибова В.В. Онтологии для разработки и генерации адаптивных пользовательских интерфейсов редакторов баз знаний / В.В. Грибова, С.В. Паршкова, Л.А. Федорищев // Онтология проектирования, 2022. – Т. 12. – № 2(44). – С. 200-217. – DOI:10.18287/2223-9537-2022-12-2-200-217.
20. CIDOC-CRM ISO page, available at: <https://www.iso.org/standard/57832.html> (accessed: 12/01/2021).
21. VK ID. Resource. – URL: <https://id.vk.com/about/business> (дата обращения: 01.07.2024).

Лисин Владислав Александрович. Аспирант Новосибирского Государственного Университета, программист второй категории Института Систем Информатики СО РАН. AuthorID: 1170641, SPIN: 3572-8093.

Сидорова Елена Анатольевна. К.ф.-м.н., старший научный сотрудник лаборатории искусственного интеллекта Института систем информатики им. А.П. Ершова СО РАН (Новосибирск), доцент кафедры программирования и кафедры систем информатики Новосибирского государственного университета. Основные направления исследований связаны с компьютерной лингвистикой, анализом аргументации, онтологическим инжинирингом и разработкой интеллектуальных систем. Author ID: 146000, SPIN: 3674-4380, ORCID: 0000-0001-8731-3058, lsidorova@iis.nsk.su, 630090, Россия, г. Новосибирск, проспект Академика Лаврентьева, 6.

UDC 004.82

DOI:10.25729/ESI.2025.38.2.015

Model of the information portal designer based on the ontologies of the subject area and resources

Vladislav A. Lisin¹, Elena A. Sidorova²

¹Novosibirsk State University,

Novosibirsk, Russia, vladlisin2@gmail.com

²Institute of Systems Informatics A.P. Ershov SB RAS,

Novosibirsk, Russia

Abstract. The article presents an approach to developing a constructor of information knowledge portals based on an ontological description of information. The applied methodology is for constructing knowledge portals based on basic ontologies and ontological design patterns. The constructor of knowledge portals should provide end-to-end technology for creating and filling a specialized Internet resource, as well as a number of intelligent services, such as an ontology editor for modeling the subject area and uploaded resources, a tool for semantic markup of texts based on an ontology and a customizable interface. The peculiarity of the proposed approach is a) the use of basic ontologies of the scientific subject area and resources as a foundation for developing specialized ontologies, b) the use of ontologies to implement the key functionality of the software associated with setting up data visualization, navigation and intelligent search on the target resource, c) putting focus on specialists who do not have experience in programming or web development. This paper proposes a formal model of a constructor of information portals and a model of an information resource created using this designer, and describes important

technical solutions to support the end-to-end technology. The created basic components of the technology are demonstrated using the example of creating a simple information system.

Keywords: ontology editor, resource ontology, page editor, knowledge portal, multimedia

References

1. Gangemi A., Presutti V. Ontology design patterns. Handbook on ontologies, Springer, 2009, pp. 221-243, DOI: 10.1007/978-3-540-92673-3_10.
2. Zagorulko Yu.A., Borovikova O.I., Zagorulko G.B. Application of ontological design patterns in the development of ontologies of scientific subject areas [Primenenie patternov ontologicheskogo proektirovaniia pri razrabotke ontologii` nauchny`kh predmetny`kh oblasti]. Analitika i upravleniye dannymi v oblastiakh s intensivnym ispol'zovaniyem dannykh: sbornik nauchnykh trudov XIX Mezhdunarodnoy konferentsii DAMDID/RCDL'2017 [Analytics and data management in data-intensive areas: collection of scientific papers of the XIX International Conference DAMDID/RCDL'2017], Moscow, FRC IU RAS, 2017, pp. 332-340.
3. Lomov P.A. Application of ontological design patterns for creation and use of ontologies within the integrated knowledge space [Primenenie patternov ontologicheskogo proektirovaniia dlia sozdaniia i ispol'zovaniia ontologii` v ramkakh integrirovannogo prostranstva znaniia]. Ontologiya proyektirovaniya [Ontology of design], 2015, vol. 5, no. 2(16), pp. 233-245.
4. Drozhdin V.V., Zhukov M.V. Expert-oriented approach to the development of application software [E`kspertno-orientirovanny`i` podhod k razrabotke prikladnogo programmnoho obespecheniia]. Problemy informatiki v obrazovanii, upravlenii, ekonomike i tekhnike: Sb. statey IX Mezhdunar. nauchno-tekhn. konf [Problems of informatics in education, management, economics and technology: Collection of articles of the IX International scientific and technical conf.], Penza, PDZ, 2009, pp. 131-135.
5. Alobaid A., Garijo D., Poveda-Villalón M., et al. Automating ontology engineering support activities with OnToology. Journal of Web semantics, 2019, no. 57, pp. 100472, DOI:10.1016/j.websem.2018.09.003.
6. Isotani S., Bittencourt I.I., Barbosa E.F., et al. Ontology driven software engineering: a review of challenges and opportunities. IEEE Latin America transactions, 2015, no. 13(3), pp. 863-869, DOI: 10.1109/TLA.2015.7069116.
7. Khoroshevsky VF. Ontology Driven Software Engineering: Models, Methods, Implementations [Proektirovanie sistem programmnoho obespecheniia pod upravleniem ontologii`: modeli, metody`, realizatsii]. Ontologiya proyektirovaniya [Ontology of design], 2019, no. 9(4), pp. 429-448.
8. Zagorulko Y.A., Zagorulko G.B. Experience of using content patterns in the development of Ontologies of scientific subject areas. Pattern recognition and image analysis, 2023, vol. 33, iss. 3, pp. 560-567, DOI:10.1134/S1054661823030513.
9. Zagorulko Y.A., Zagorulko G.B., Sery A.S. et al. A software package for the development of intelligent decision support systems in poorly formalized areas [Programmny`i` kompleks dlia razrabotki intellektual`ny`kh sistem podderzhki priniatiia reshenii` v slaboformalizovanny`kh oblastiakh]. Nauka i tekhnologii Sibiri. Sibirskaya industriya informatsionnykh sistem [Science and technology of Siberia, Siberian information systems industry], 2023, no.10, pp. 34-38.
10. Sapetina A., Kulikov I., Zagorulko G., et al. Constructing an expert system for solving astrophysical problems based on the ontological approach. Parallel computational technologies, communications in computer and information science, Springer, 2022. vol. 1618, pp. 30-42, DOI:10.1007/978-3-031-11623-0_3.
11. Zagorulko Y., Borovikova O., Zagorulko G. Knowledge portal on computational linguistics: content-based multilingual access to linguistic information resources. Selected topics in Applied computer science. Proceedings of the 10th WSEAS International Conference on Applied Computer Science (ACS'10). Hamido Fujita, Jun Sasaki (Eds.), WSEAS Press, 2010, pp. 255-262.
12. Sery A.S., Grinevich A.A., Lisin V.A. An approach to building an information research environment for analyzing scientific materials and creating annotated corpora [Podhod k postroeniui informatcionnoi` issledovatel'skoi` sredi` dlia analiza nauchny`kh materialov i sozdaniia annotirovanny`kh korpusov]. Informatsionnyye i matematicheskiye tekhnologii v nauke i upravlenii [Information and mathematical technologies in science and management], 2021, no. 4 (24), pp. 68-76.
13. CIDOC Conceptual reference model, available at: <http://www.cidoc-crm.org/> (accessed: 12/01/2021).
14. Lisin V.A., Sidorova E.A. Web platform for folklore research based on the ontology of subject areas [Veb-platforma dlia fol'clorny`kh issledovaniia na osnove ontologii predmetny`kh oblasti]. Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Informatsionnyye tekhnologii [Bulletin of the Novosibirsk state university. Series: information technologies], 2021, vol. 19, no. 2, pp. 53-64.
15. Lisin V.A., Sery A.S., Sidorova E.A. Model for representing the ontology of subject areas based on graph databases [Model` predstavleniia ontologii predmetny`kh oblasti` na osnove grafovy`kh baz danny`k]. Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Informatsionnyye tekhnologii [Bulletin of the Novosibirsk

- state university. Series: information technologies], 2022, vol. 20, no. 4, pp. 24-38, DOI: 10.25205/1818-7900-2022-20-4-24-38.
16. Jesús Barrasa RDF Triple stores vs. Labeled property graphs: what's the difference? available at: <https://neo4j.com/blog/rdf-triple-store-vs-labeled-propertygraph-difference> (accessed: 01/13/2023).
 17. Grinevich A., Sery A. Analysis of cultural universals of folklore of the peoples of Siberia and the Far East [Analiz kul'turny`kh universalii` fol`clora narodov Sibiri i Dal`nego Vostoka]. Materialy XXIII Mezhdunarodnoy konferentsii po analitike i upravleniyu dannymi v informatsionno yemkikh oblastiakh (DAMDID/RCDL 2021) [Proceedings of the XXIII International Conference on Analytics and Data Management in Information-Intensive Areas (DAMDID/RCDL 2021)], Moscow, 2021, no. 3036, pp. 387-401.
 18. Zagorulko Yu.A. Semantic technology for developing intelligent systems aimed at subject area experts [Semanticheskaiia tekhnologiia razrabotki intellektual`ny`kh sistem, orientirovannaia na e`kspertov predmetnoi` oblasti]. Ontologiya proyektirovaniya [Ontology of design], 2015, no. 5(1), pp. 30-46.
 19. Gribova V.V., Parshkova S.V., Fedorishchev L.A. Ontologies for developing and generating adaptive user interfaces of knowledge base editors [Ontologii dlia razrabotki i generatsii adaptivny`kh pol`zovatel`skikh interfei`sov redaktorov baz znaniï]. Ontologiya proyektirovaniya [Ontology of design], 2022, no. 12(2), pp. 200-217.
 20. CIDOC-CRM ISO page, available at: <https://www.iso.org/standard/57832.html> (accessed: 12/01/2021).
 21. VK ID, available at: <https://id.vk.com/about/business> (accessed: 07/01/2024).

Lisin Vladislav Aleksandrovich. PhD student at Novosibirsk State University, programmer of the second category at Institute of Informatics Systems SB RAS, (Novosibirsk, Russia). AuthorID: 1170641. SPIN: 3572-8093.

Sidorova Elena Anatolievna. PhD, Senior researcher of the Laboratory of Artificial Intelligence at the A.P. Ershov Institute of informatics systems SB RAS, associate professor at Novosibirsk state university. The research areas are related to Computational Linguistics, Intelligent System Development, Knowledge and Ontology Engineering. AuthorID: 146000, SPIN: 3674-4380, ORCID: 0000-0001-8731-3058, Isidorova@iis.nsk.su, 630090, Russia, Novosibirsk, Academician Lavrentiev Avenue, 6.

Статья поступила в редакцию 17.12.2024; одобрена после рецензирования 20.03.2025; принята к публикации 11.04.2025.

The article was submitted 12/17/2024; approved after reviewing 03/20/2025; accepted for publication 04/11/2025.