

в различных учебных дисциплинах, связанных с охраной окружающей среды и экологией человека и рекомендованы к внедрению в лесохозяйственных организациях и при подготовке специалистов в области лесоведения в университетах России.

зайственных организациях и при подготовке специалистов в области лесоведения в университетах России.

СПИСОК ЛИТЕРАТУРЫ

1. RINNTech // Products and services for tree and wood analysis. 2010. URL: <http://www.rinntech.de/> (дата обращения: 27.08.2010).
2. GPSS // Имитационное моделирование систем. 2010. URL: <http://www.gpss.ru/> (дата обращения: 27.08.2010).
3. Ботыгин И.А., Попов В.Н., Тартаковский В.А. Математические модели в задачах обработки дендрэкологических данных. Ч. I // Известия Томского политехнического университета. – 2011. – Т. 319. – № 5. – С. 118–122.
4. Ботыгин И.А., Попов В.Н., Тартаковский В.А. Математические модели в задачах обработки дендрэкологических данных. Ч. II // Известия Томского политехнического университета. – 2011. – Т. 319. – № 5. – С. 123–125.

Поступила 29.06.2011 г.

УДК 004.4:004.89

ИСПОЛЬЗОВАНИЕ ОНТОЛОГИИ В ЭЛЕКТРОННЫХ БИБЛИОТЕКАХ

Ле Хоай, А.Ф. Тузовский

Томский политехнический университет
E-mail: lehotomsk@yahoo.com

Рассматривается использование онтологий в семантических электронных библиотеках, дается их определение и назначение. Анализируются виды онтологий таких библиотек, в том числе системы организации знаний и структурная таксономия. Обосновывается вариант набора онтологий для разработки семантических электронных библиотек.

Ключевые слова:

Онтология, контрольный словарь электронных ресурсов, семантические технологии, электронная библиотека, семантическая электронная библиотека.

Key words:

Ontology, glossary of electronic resources, semantic technology, electronic library, semantic digital library.

Под электронными библиотеками (ЭБ) понимаются информационные системы, позволяющие автоматизировать работу пользователей с электронными ресурсами (ЭР), такими, как документы, изображения, аудио-, и видеофайлы и т. д. С появлением семантических технологий (СТ), предоставляющих средства работы с семантикой документов, возникла возможность разработки подходов к автоматизации работы с этими ресурсами на новом уровне. Разработка семантических электронных библиотек (СЭБ) представляет собой решение комплекса задач, целью которых являются повышение возможностей взаимодействия с пользователями и расширение функциональности ЭБ, особенно в поиске данных. Многие ЭР содержат метаданные, в том числе документы включают данные об авторе, издании, дате создания и т. д. Такие метаданные часто хранятся в виде XML файлов, позволяющих выполнять универсальное описание ЭР. Язык XML позволяет описывать только структуру объектов, а не их семантику. В свою очередь, СТ представляют возможность описывать семантику ЭР (аннотировать их) и выполнять программную обработку таких метаданных.

Для описания семантики метаданных необходимо использовать онтологические модели (онто-

логии), определяющие наборы понятий предметной области и их взаимосвязи. Для работы с онтологиями в СТ имеются специальные языки для описания семантики (OWL, RDFS, RDF) [1–3] и запросов к семантическим данным (SPARQL) [4], а также набор инструментов редактирования и работы с семантическими данными (Protégé, Sesame, Jena и т. д.) [5–7].

При разработке СЭБ с использованием СТ одной из наиболее важных и сложных задач является разработка онтологий, описывающих области знаний, с которыми связано функционирование ЭБ и содержание имеющихся в них ресурсов. В данной статье рассматриваются особенности и проблемы, связанные с задачей построения онтологий для СЭБ, и обосновывается базовый набор таких онтологий.

1. Использование онтологии в СЭБ

В области искусственного интеллекта под онтологией понимается специальная система понятий и взаимосвязей между ними, описывающая определенную предметную область. Содержание понятий определяется с помощью концептов. Формально в онтологии концепт отождествляется с объектом (классом), имеющим связи с другими класса-

ми. Класс определяется как множество экземпляров с общими свойствами и содержит описания собственно экземпляров и их свойств [8].

Для обеспечения возможности программной работы с семантикой был разработан язык RDF для описания высказываний, а также язык OWL для описания схемы отношений между концептами (классами) и их атрибутами (свойствами) [9]. С помощью этих языков и стандартов программа понимает данные и может выполнять логические выводы.

На основе различных метаданных СЭБ интегрирует разнородную информацию, например, ресурсы, пользовательские профили, закладки и таксономии. В связи с этим в СЭБ онтологии играют основную роль для решения проблем, вызванных структурными различиями существующих систем и семантическими различиями стандартов метаданных.

2. Типы онтологий СЭБ

Из анализа ранее выполненных исследовательских проектов по разработке СЭБ (например, JeromeDL [10], BRICKS [11]) можно определить три типа онтологий, используемых в библиотеках: 1) библиографические; 2) структурные; 3) пользовательские.

2.1. Библиографические онтологии

Для хранения и использования информации о ресурсах библиотеки создаются библиографические форматы (стандарты). Примером набора таких стандартов является СИБИД (система стандартов по информации, библиотечному и издательскому делу), которая основана на аналогических международных стандартах, устанавливающих термины и определения в области информационной деятельности, библиотечного дела и библиографии.

Существует ряд форматов описания библиографической информации, например, MARC (MARC21, RUSMARC, UNIMARC), ONIX XML, BibTEX, Dublin Core, каждый из которых имеет свою область применения. Библиотеки обычно используют форматы семейства MARC; исследователи хранят информацию о публикациях в виде BibTEX; а пользователи Интернет в виде Dublin Core. Данные стандарты определяют набор резервированных полей и позволяют хранить данные в виде обычного текста, но компьютерные программы не могут осуществлять интеллектуальную обработку этих значений. В связи с этим возникает необходимость использования специальных полей с семантическими описаниями. Набор таких полей формирует основу онтологии, позволяющей описывать понятия и отношения предметной области.

Онтология дополняет существующие форматы семантикой и позволяет осуществлять интеграцию данных для них [10].

Онтология MarcOnt была разработана в качестве однородного языка описания библиографиче-

ских данных путем объединения таких существующих онтологий, как BibTeX-RDF, Dublin Core и MarcRDF. Одной из причин разработки MarcOnt [12] явилась сложность языка MARC. Онтология MarcOnt использовалась в качестве базовой в соответствующем проекте MarcOnt, разрабатывающем подход к работе с описаниями метаданных. В проекте был также разработан набор программных средств Mediation Services, предоставляющих возможности выполнения преобразования между разнородными форматами описаний ЭР.

Онтология SWRC (Semantic Web for Research Communities) [13] была разработана институтом AIFB для описания простых библиографических записей и дополнительной информации об исследовательских сообществах. В терминах библиографической выразительности онтология похожа на BibTeX и Dublin Core. В исследовательском проекте Bibster онтология SWRC объединена с онтологическим представлением ACM иерархии тем.

Библиографическая онтология ViVo предоставляет главные концепты и атрибуты для описания цитат и библиографических ссылок. Онтология ViVo использует такие словари (онтологии), как Dublin Core и FOAF. Она позволяет описать ресурсы (документы, книги, статьи), их коллекции и другие объекты. Например, эта онтология включает такое понятие, как Event для описания события, связанного с ресурсами.

Из рассмотрения данной структуры можно отметить, что онтология ViVo является сочетанием структурной и библиографической онтологий, из которых структурная онтология может быть выделена отдельно.

2.2. Структурная онтология

Современные электронные библиотеки хранят не только библиографические метаданные ЭР, но и представление их содержания. В зависимости от их типов, содержание ресурса обычно следует некоторой структуре, которую можно занести в метаданные описания. Например, книги, хранимые в библиотеке, обычно делятся на главы, и для каждой главы можно предоставить индивидуальное описание и хранить информацию об отношениях данной главы с другими главами.

Одним из примеров такой онтологии является структурная онтология библиотеки JeromeDL. Кроме структурного описания ресурсов, в этой онтологии можно подробнее описать статус каждого ресурса, данные о дате поступления в библиотеку и другую информацию, необходимую для управления работой системы с данным ресурсом [14].

2.3. Пользовательская онтология

Данный тип онтологий предназначен для описания профилей пользователей библиотеки, включая их интересы в определенных ресурсах и другую информацию. Существуют онтологии FOAF [15] и FOAFRealm [16], которые позволяют описать профили пользователей. Онтология FOAFRealm

является расширенным вариантом FOAF, с ее помощью можно определить насколько пользователи хорошо знают друг друга, а также она позволяет хранить зашифрованные данные. Такая онтология использовалась в исследовательском проекте JermDL, и для работы с ней был разработан набор программных средств. В этих онтологиях можно добавлять и другие атрибуты для целей решения собственных задач.

Следует отметить, что пользователи библиотеки могут иметь заинтересованность в определенных ресурсах или темах. Подробнее вопрос будет рассмотрен при описании задач разработки СЭБ.

2.4. Другие онтологии в СЭБ

Кроме основных рассмотренных онтологий, в СЭБ используются и другие онтологии, с помощью которых создаются структурные контрольные словари, важные для категоризации библиотечных ресурсов. Структурные контрольные словари представляют собой тезаурусы, схемы классификации, таксономии, выражаемые в виде схемы концептов (SKOS/RDF). Существуют несколько видов контрольных словарей [17]. Чаще всего используемыми словарями для классификации ресурсов являются таксономическая классификация ACM (для компьютерных наук), DDC (десятичная классификация Дьюи) и UDC (универсальная десятичная классификация).

Применение семантических технологий для этих словарей предоставляет возможность аннотировать ресурсы и решить задачи каталогизации и категоризации. Существуют варианты кодирования этих словарей в формат SKOS/RDF для применения в СЭБ [18]. Так как эти варианты применяются только для аннотирования ресурсов на английском языке, то при использовании русского языка необходимо изменить их кодирование в формате SKOS/RDF. Кроме этого, требуется внести в них и ряд дополнений. Например, для диссертационных работ необходимо иметь возможность указывать код специальностей ВАК [19]. Для этого необходимо разработать и включить в SKOS/RDF контрольный словарь, в котором каждый код является экземпляром класса концепта и имеет название – этикетку (на русском или другом языке).

В СЭБ базовыми являются онтологии, на основе которых решаются основные задачи их функционирования.

3. Основные задачи функционирования семантических ЭБ

- **Ведение онтологической базы знаний:** задача может быть решена только после создания всех онтологий СЭБ. Для решения задачи требуется разработать программные инструменты для создания базы знаний.
- **Ведение иерархии каталогов:** задача представляет собой создание иерархической структуры

электронных документов с использованием какой-либо категоризации. Для каждой категоризации должна быть разработана возможность представления ЭР в виде иерархии каталогов.

- **Аннотирования ресурсов:** задача аннотирования ЭР должна быть решена путем представления подходящих пользовательских интерфейсов для внесения требуемых данных в описания ЭР.
- **Категоризация:** задача тесно связана с задачей ведения иерархии каталогов, суть задачи заключается в создании специальных словарей для распределения ресурсов по типам, видам и особенностям.
- **Просмотр:** предоставляется возможность удобного перехода между данными электронных документов (автор, издание, ключевые слова и т. д.).
- **Поиск:** это типичная задача каждой системы управления данными. Новые технологии и СТ должны способствовать более точному и полному поиску документов.
- **Формирование профилей пользователей:** задача включает заполнение профилей пользователей данными с использованием удобного пользовательского интерфейса.
- **Рекомендация:** задача решается с использованием учетных данных профилей пользователей. Интересы пользователей являются ключевыми для рекомендации им соответствующих документов.

Выше перечисленные задачи должны учитываться при создании онтологий для СЭБ, а также необходимо использовать некоторые методологии [20–22].

4. Предложенный вариант базовых онтологий СЭБ

4.1. Пользовательская онтология

В качестве онтологии пользователей рекомендуется модифицированная онтология FOAF для описания пользовательских профилей. Схема предложенной онтологии, построенная с помощью редактора Protégé, приведена на рис. 1. Базовым классом онтологии является класс «Агент», одним из подклассов которого – класс «Человек». В связи с этим все связи класса «Агент» с другими классами можно рассматривать как связи класса «Человек».

Полное описание онтологии FOAF приведено в [23], а далее описаны только новые добавленные, выделенные подчеркиванием, классы (концепты) и атрибуты:

- **Домен интереса** «URI: <http://www.w3.org/2004/02/skos/core#Concept>»: включение класса объясняется тем, что в СЭБ должны храниться интересы пользователя для рекомендации ему новых документов. Так как каждый документ в СЭБ относится к какой-либо области и если она интересует пользователя, то все новые поступления документов будут рекомендованы ему, и он будет проинформирован об этих документах.

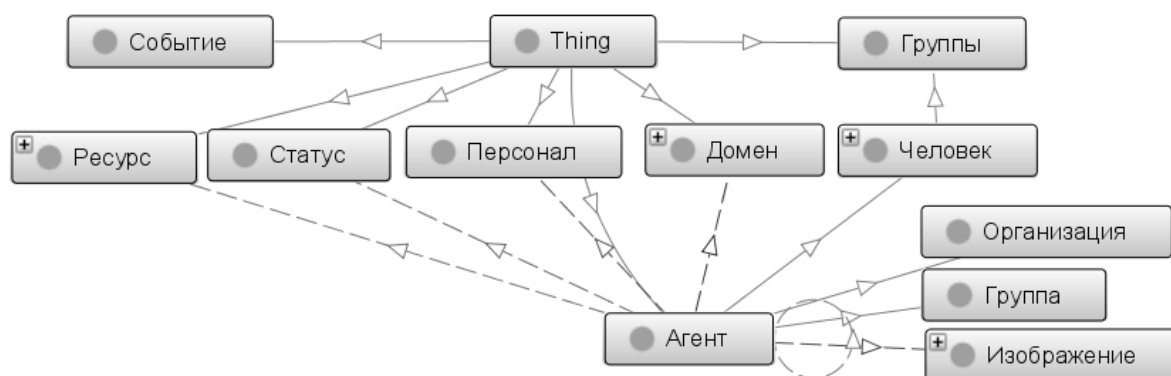


Рис. 1. Схема онтологии пользовательских профилей

- **Персонал библиотеки** «URI: <http://semdl.tpu.ru/ontology/structure/1.0/LibraryStaff>»: класс имеет только три экземпляра (**Admin**, **Editor**, **User**). Каждый зарегистрированный пользователь будет являться одним из них и иметь соответствующие права.
- **Группы** «URI: <http://semdl.tpu.ru/ontology/structure/1.0/Groups>»: Пользователь библиотеки может находиться в дружеских отношениях с другими пользователями, которые группируются по пользовательским целям для обмена информацией об электронных документах.
- **Статус** «URI: <http://semdl.tpu.ru/ontology/structure/1.0/StatusType>»: класс обладает тремя индивидуумами, такими как **Online**, **Busy**, **Offline** для обозначения статуса пользователя в системе.
- **Событие** «URI: <http://purl.org/NET/c4dm/event.owl#Event>»: класс представляет событие, которое может произойти, и о котором система будет информировать пользователей.
- **Ресурсы**: «URI: <http://www.marcont.org/ontology/2.0#Resource>»: класс включен в онтологию, т. к. в онтологии ресурсов он является основным и суперклассом других классов (книги, статьи, доклады и т. д.). В данной онтологии все атрибуты связей остаются прежними, а основное внимание было уделено таким атрибутам, как «**интересуется-Темой**» (`topic_interest`) и «**интересует**» (`interest`):
 - **ИнтересуетсяТемой** «URI: http://xmlns.com/foaf/0.1/topic_interest, Domains: **Агент**, Ranges: **Событие**, **Домен** **Интереса**»: заполнение данной связи пользователем позволит системе рекомендовать и информировать его о новом поступлении интересующих его документов.
 - **Интересует** «URI: <http://xmlns.com/foaf/0.1/interest>, Domains: **Агент**, Ranges: **Thing**»: для избавления от путаницы при выполнении логического вывода были добавлены два атрибута (**interestAgent**, **interestResource**), которые позволяют пользователям библиотеки описывать свои интересы (издания, книги, автора и т. п.). Данная связь также позволяет рекомендовать и информировать пользователя о новом поступлении документов конкретного автора или новых комментариев к ним.

В онтологии пользователей центральным является класс «Агент–Человек», то в онтологии ресурсов или онтологии предметных областей класс «Ресурс».

4.2. Библиографическая онтология

Библиографическая онтология (или онтология ресурсов) основывается на библиотечной онтологии Marcont [24], и при решении задач СЭБ будет использоваться эта онтология с некоторыми изменениями. В данной онтологии содержатся необходимые классы и атрибуты для описания электронных документов, таких, как книга, статья, диссертация и т. д. Схема онтологии приведена на рис. 2.

Класс «Ресурс» представляет все типы электронных документов в СЭБ и имеет связь с классом «Агент–Человек», в котором описываются издание, автор, редактор, соучастник. Этот класс также связывается с классом «Домен», представляющим собой концепт системы организации знаний (SKO); концептом могут быть ключевые слова, домен области документов.

Каждый документ в СЭБ может быть одним из типов ресурсов (книга, статья, диссертация и другие), а это значит, что задачу каталогизации документов можно решить с использованием таких типов. Задача категоризации заключается в аннотировании документов по своим доменам (класс «Домен»).

4.3. Структурная онтология (онтология структуры)

Онтология позволяет описывать каждый документ метаданными о его структуре, типе, оценке пользователя, а также коллекции. Кроме этого введены новые классы, такие, как класс «Сообщение» для представления разных оповещений от системы к пользователю и класс «Тип Статуса Документа» для задания статуса документа (**IPublished**, **ISubmitted**, **IToBeSubmitted**, **IAccepted**) при его редактировании библиотечным персоналом.

Схема онтологии структуры СЭБ показана на рис. 3

В отличие от других данная онтология характеризует работу конкретной системы и повторно не может быть использована. Вариант онтологии подобного типа можно найти в [25].

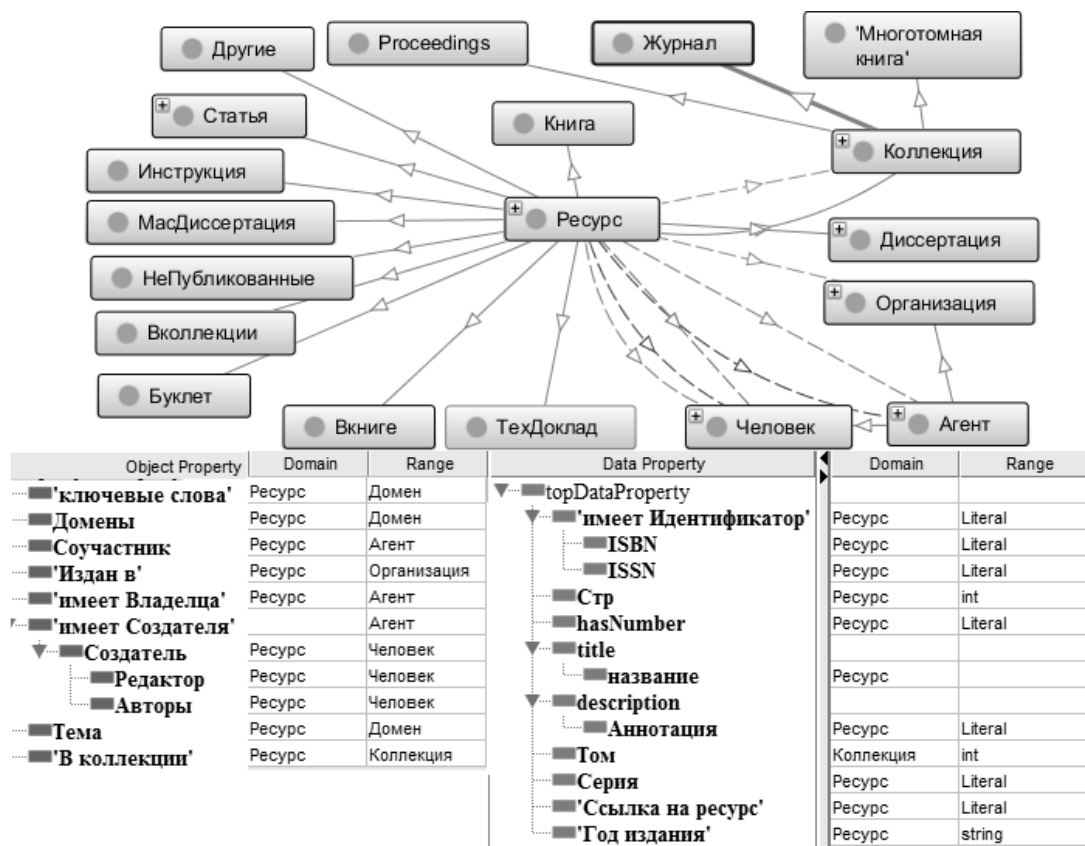


Рис. 2. Схема онтологии ресурсов (электронных документов) и их атрибуты

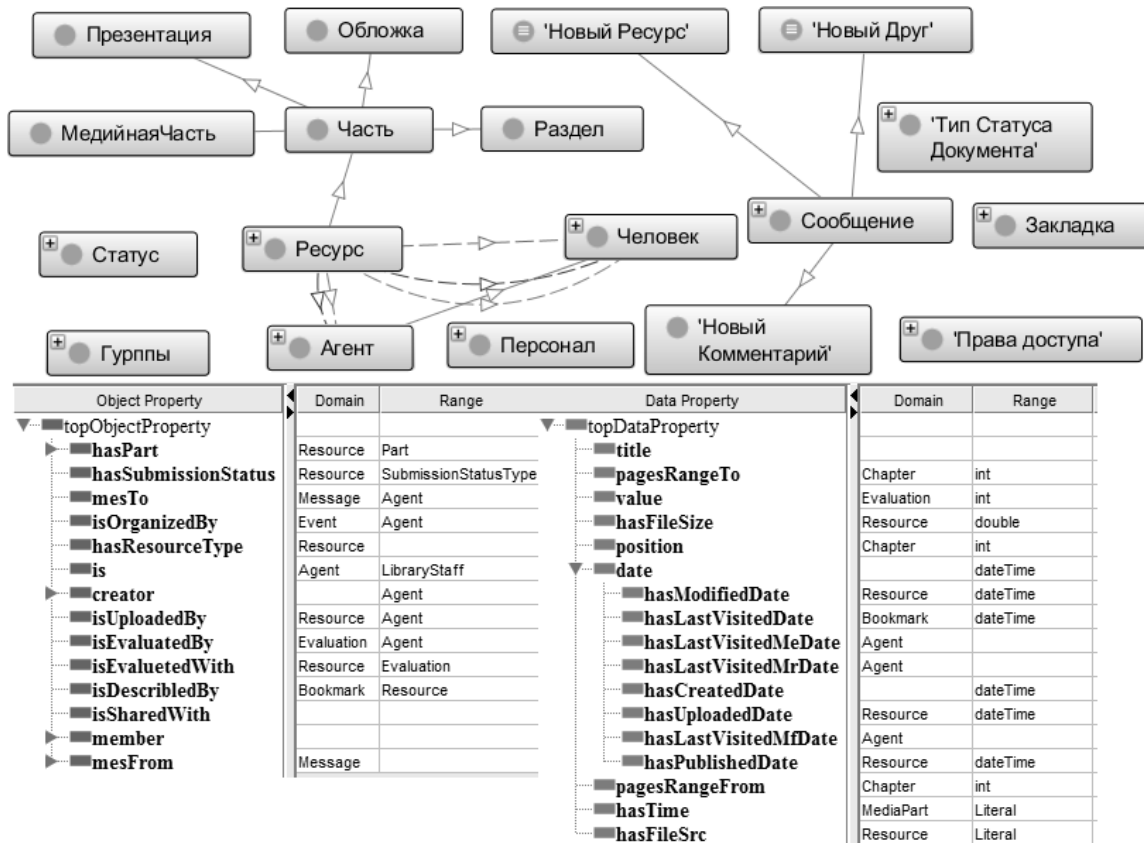


Рис. 3. Схема онтологии структуры электронных ресурсов и их атрибуты

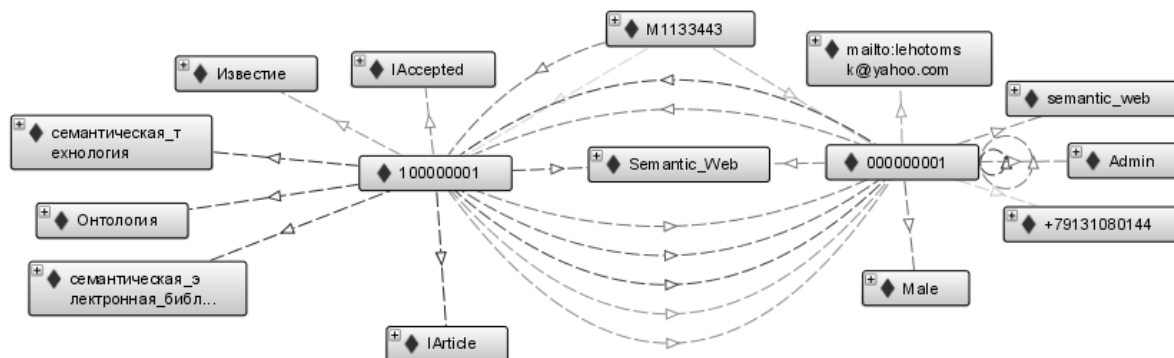


Рис. 4. Схема примера использования онтологий

4.4. Пример использования

Рассмотрим пример использования представленных выше онтологий для описания пользователя, документа и его системных данных. На рис. 4 показана схема отношений между экземплярами классов этих онтологий, где **000000001** – **Человек**, **100000001** – **Статья**.

На схеме не показаны свойства, но их можно посмотреть в виде OWL файла:

Выводы

Для создания семантических электронных библиотек требуется разрабатывать набор специальных онтологий. В качестве таких могут использоваться библиографическая, структурная и пользовательская онтологии. Их применение позволяет решать основные задачи разработки семантических электронных библиотек. Разработанный вариант онтологий будет применен для создания базы знаний семантических электронных библиотек.

```

<Person rdf:about="&users;000000001">
  <rdf:type rdf:resource="&owl;NamedIndividual"/>
  <phone rdf:resource="&FoafUser;+79131080144"/>
  < users:interestAgent rdf:resource="& users;000000001"/>
  < users:interestResource rdf:resource="&Resource;100000001"/>
  < users:gender rdf:resource="& users;Male"/>
  <mbox rdf:resource="& users;mailto:lehotomsk@yahoo.com"/>
  <topic_interest rdf:resource="&resource;Semantic_Web"/>
  <structure:is rdf:resource="&structure;Admin"/>
</Person>
<owl:NamedIndividual rdf:about="&resource;100000001">
  <rdf:type rdf:resource="http://www.marcont.org/ontology/2.0#Article"/>
  <hasPages rdf:datatype="&xsd:int">8</hasPages>
  <hasAbstract rdf:datatype="&xsd:string">В статье рассматривается вариант онтологий
  для разработки семантических электронных библиотек (СЭМ), их спецификации для решения
  основных задач в системе управления электронными документами, а также проводится изучение
  методологии создания онтологий.</hasAbstract>
  <resource:hasTitle rdf:datatype="&xsd:string">Разработка онтологий для семантических
  электронных библиотек</Resource:hasTitle>
  <hasYear>2011</hasYear>
  <hasAuthor rdf:resource="&users;000000001"/>
  <hasDomain rdf:resource="&resource;Semantic_Web"/>
  <hasPublisher rdf:resource="&resource;Известие"/>
  <hasKeyword rdf:resource="&resource;Онтология"/>
  <hasKeyword rdf:resource="&resource;семантическая_технология"/>
  <hasKeyword rdf:resource="&resource;семантическая_электронная_библиотека"/>
  <rdf:type rdf:resource="&ontology;Resource"/>
  <isUploadedBy rdf:resource="&users;000000001"/>
  <hasSubmissionStatus df:resource="http://semDl.tpu.ru/ontology/structure/1.0/IAccepted"/>
  <hasResourceType rdf:resource="http://semDl.tpu.ru/ontology/structure/1.0/IArticle"/>
</owl:NamedIndividual>
    
```

СПИСОК ЛИТЕРАТУРЫ

1. OWL Web Ontology Language Overview. 2011. URL: <http://www.w3.org/TR/owl-features/> (дата обращения: 20.04.2011).
2. RDF Vocabulary Description Language 1.0: RDF Schema. 2011. URL: <http://www.w3.org/TR/rdf-schema/> (дата обращения: 20.04.2011).
3. Resource Description Framework (RDF). 2011. URL: <http://www.w3.org/RDF/> (дата обращения: 20.04.2011).
4. SPARQL Query Language for RDF. 2011. URL: <http://www.w3.org/TR/rdf-sparql-query/> (дата обращения: 20.04.2011).
5. Welcome to the Protege wiki. 2011. URL: http://protegewiki.stanford.edu/wiki/Main_Page (дата обращения: 20.04.2011).
6. Java RDF Framework and Database System. 2011. URL: <http://sourceforge.net/projects/sesame/> (дата обращения: 20.04.2011).
7. Jena – A Semantic Web Framework for Java. 2011. URL: <http://jena.sourceforge.net/> (дата обращения: 20.04.011).
8. Кузнецов О.П., Суховеров В.С., Шипилина Л.Б. Онтология как систематизация научных знаний: структура, семантика, задачи // Технические и программные средства систем управления, контроля и измерения: Труды II Росс. конф. с международным участием. – М., 2010. – С. 762–773.
9. RIF RDF and OWL Compatibility. 2011. URL: <http://www.w3.org/TR/rif-rdf-owl/> (дата обращения: 20.04.2011).
10. Kruk S.R., Decker S., Zieborak L. JeromeDL – Adding Semantic Web Technologies to Digital Libraries // Lecture Notes in Computer Science. – 2005. – V. 3558/2005. – P. 716–725.
11. BRICKS Project: Building Resources for Integrated Cultural Knowledge Services (IST 507457). 2011. URL: <http://www.brickcommunity.org> (дата обращения: 20.04.2011).
12. Kruk S.R., Synak M., Kerstin Z. MarcOnt – Integration Ontology for Bibliographic Description Formats // International Conference on Dublin Core and Metadata Applications: Proc. – Madrid, Spain, 2005. – P. 231–234.
13. York S., Stephan B., Peter H., Jens H., Daniel O. The SWRC ontology – semantic web for research communities // 12th Portuguese Conference on Artificial Intelligence – Progress in Artificial Intelligence: Proc. – 2005. – V. 3803 of LNCS. – Covilha, Portugal, DEC Springer 2005. – P. 218–231.
14. The Role of Ontologies in Semantic Digital Libraries. 2011. URL: <http://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2006/papers/paper-13.pdf> (дата обращения: 23.04.2011).
15. OWL Web Ontology Language. 2011. URL: <http://www.w3.org/TR/owl-guide/> (дата обращения: 23.04.2011).
16. FOAF-Realm – control your friends' access to resources. 2011. URL: http://www.w3.org/2001/sw/Europe/events/foaf-galway/papers/fp/foaf_realm/ (дата обращения: 23.04.2011).
17. Linking Public Vocabularies. 2011. URL: <http://www.openvocabulary.info/vocabularies/> (дата обращения: 23.04.2011).
18. JOnto project. 2011. URL: <http://jonto.sourceforge.net/> (дата обращения: 25.04.2011).
19. Паспорта специальностей ВАК. 2011. URL: <http://teacode.com/online/vak/> (дата обращения: 25.04.2011).
20. Dean A., James H. Semantic web for the working ontologist modeling in RDF, RDFS and OWL. – Morgan Kaufmann, 2008. – 352 p.
21. Yu Liyang. A Developer's Guide to the Semantic Web. – Springer, 2011. – 628 p.
22. Ontology editor. 2011. URL: http://en.wikipedia.org/wiki/Ontology_editor (дата обращения: 25.04.2011).
23. FOAF Vocabulary Specification 0.98. 2011. URL: <http://xmlns.com/foaf/spec/> (дата обращения: 26.04.2011).
24. Marcont Vocabulary Specification 2.0. 2011. URL: <http://www.marcont.org/ontology/2.0> (дата обращения: 26.04.2011).
25. JeromeDL Ontology Specification. 2011. URL: <http://www.jeromedl.org/structure/2.1> (дата обращения: 27.04.2011).

Поступила 27.01.2012 г.