

Организация содержательного доступа к гуманитарным информационным ресурсам на основе онтологий*

© Загорулько Ю.А.

Институт систем
информатики имени А.П.
Ершова СО РАН
zagor@iis.nsk.su

Боровикова О.И.

Институт систем
информатики имени А.П.
Ершова СО РАН
olesya@iis.nsk.su

Загорулько Г.Б.

Институт систем
информатики имени А.П.
Ершова СО РАН
gal@iis.nsk.su

Аннотация

В докладе представлен подход к организации содержательного доступа к гуманитарным информационным ресурсам на основе онтологий. Онтология составляет информационную основу Интернет-портала знаний, который должен обеспечить интеграцию и систематизацию научных знаний и информационных ресурсов определенной тематики, а также содержательный доступ к ним из любой «точки» Интернет-пространства. По онтологии автоматически строятся схема внутренней базы данных портала и формы для ее заполнения, организуется навигация по информационному пространству портала и обеспечивается формулирование поисковых запросов в терминах предметной области портала знаний. Разделение онтологии портала на предметно-независимые и предметные онтологии делает портал настраиваемым практически на любую область научных знаний.

1 Введение

В настоящее время в сети Интернет представлен большой объем знаний и информационных ресурсов гуманитарной направленности. Однако эти ресурсы не достаточно систематизированы, т.е. практически случайным образом распределены по каталогам, электронным архивам или размещены на отдельных сайтах как гуманитарной, так и технической направленности, что значительно затрудняет их поиск и использование.

Для решения этой проблемы нами была предложена концепция и архитектура специализированного Интернет-портала знаний [3, 22], который должен обеспечить интеграцию и систематизацию научных знаний и

информационных ресурсов определенной тематики, а также содержательный доступ к ним из любой «точки» Интернет-пространства. Согласно этой концепции, портал знаний должен обеспечивать доступ не только к собственным информационным ресурсам, но и поддерживать навигацию по заранее размеченным (проиндексированным) ресурсам, размещенным в сети Интернет.

В качестве информационной модели портала знаний используется онтология, что позволило не только обеспечить его основную функциональность, но и сделало его настраиваемым на любую область научных знаний.

2 Онтология портала знаний

Основываясь на идеях, изложенных в [6, 13, 15, 20], под онтологией мы будем понимать точное подробное описание (модель) некоторой части мира применительно к конкретной области интересов. В контексте данной работы онтология будет представлять собой описание некоторой гуманитарной дисциплины (науки) и научной деятельности, связанной с этой дисциплиной.

Формально онтология портала знаний может быть описана семеркой вида:

$$O = \langle C, A, R_C, T, D, R_A, F \rangle, \text{ где}$$

C – множество классов, описывающих понятия некоторой предметной или проблемной области;

A – множество атрибутов, описывающих свойства понятий и отношений;

$R_C = \{r_c \mid r_c \subseteq C \times C\}$ – множество отношений, заданных на классах (понятиях);

T – множество стандартных типов значений атрибутов (*string, integer, real, date*);

D – множество доменов (множеств значений стандартного типа *string*);

$R_A = R_{AT} \cup R_{AD}$, где $R_{AT} \subseteq A \times T$ – отношение, связывающее атрибуты и типы данных, из которых они могут принимать свои значения, $R_{AD} \subseteq A \times D$ – отношение, определяющее для каждого атрибута его дискретное множество значений (домен);

F – множество ограничений на значения атрибутов понятий и отношений.

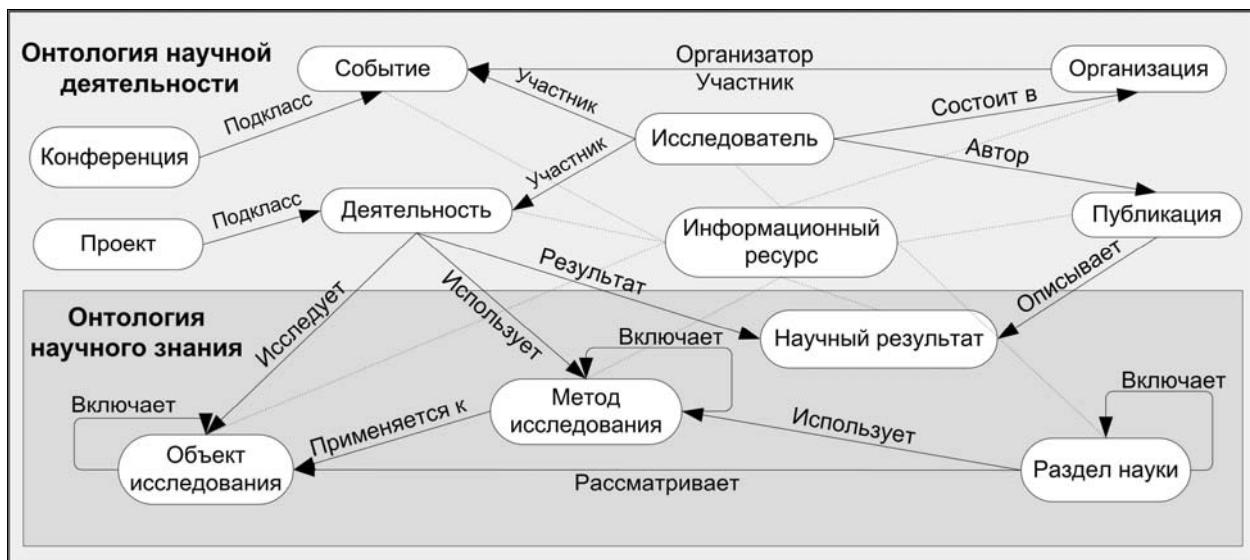


Рис.1. Базовые онтологии портала знаний

С содержательной точки зрения, онтология портала служит для представления понятий, необходимых для описания как научной деятельности и научного знания в целом, так и конкретной области знаний, в частности.

Для упрощения настройки портала на выбранную область знаний онтология портала разделяется на предметно-независимые (базовые) онтологии и онтологию предметной области (предметную онтологию).

В качестве базовых выбраны онтологии научной деятельности и научного знания (Рис.1), которые не зависят от предметной области портала.

Онтология научной деятельности построена на основе онтологии, предложенной в [8] для описания научно-исследовательских проектов, и, фактически, является онтологией верхнего уровня. Она включает базовые классы понятий, относящиеся к организации научной и исследовательской деятельности, такие как *Исследователь*, *Организация*, *Событие*, *Деятельность*, *Публикация* и др. В эту онтологию также включен класс *Информационный ресурс*, который служит для описания информационных ресурсов, представленных в сети Интернет.

Онтология научного знания фиксирует основные содержательные структуры, используемые для построения предметных онтологий. В частности, эта онтология содержит метапонятия, задающие структуры для описания понятий конкретной области знаний, такие как *Раздел науки*, *Метод исследования*, *Объект исследования*, *Научный результат*.

Понятия базовых онтологий связаны между собой ассоциативными отношениями, выбор которых осуществлялся не только исходя из полноты представления проблемной и предметной областей портала, но и с учетом удобства навигации по его информационному пространству и поиска информации.

Наиболее важными ассоциативными отношениями являются:

- «научное направление» – связывает события, публикации, организации, исследователей и информационные ресурсы с разделами науки;
- «описывает» – задает связь публикации с научным результатом, объектом или методом исследования;
- «использует» – связывает метод исследования с деятельностью, исследователем или разделом науки;
- «применяется к» – связывает метод исследования с объектом исследования;
- «исследует» – сопоставляет какую-либо деятельность или раздел науки с объектом исследования;
- «результат» – связывает научный результат с деятельностью;
- «ресурс» – связывает информационный ресурс с событиями, публикациями, исследователями, методами и объектами исследования.

Понятия предметной онтологии, описывая конкретную научную дисциплину, в то же время являются реализациями метапонятий онтологии научного знания и могут быть упорядочены в иерархию «общее-частное». При построении такой иерархии разработчикам необходимо учитывать не только то, насколько она полно описывает предметную область и связанные с ней информационные ресурсы, но и насколько удобно пользователю с ее помощью осуществлять навигацию по информационному пространству портала и вести содержательный поиск.

Например, разрабатываемая нами онтология компьютерной лингвистики (Рис.2) включает четыре базовых иерархии: иерархию разделов, иерархию объектов, иерархию методов исследования и иерархию научных результатов компьютерной лингвистики.

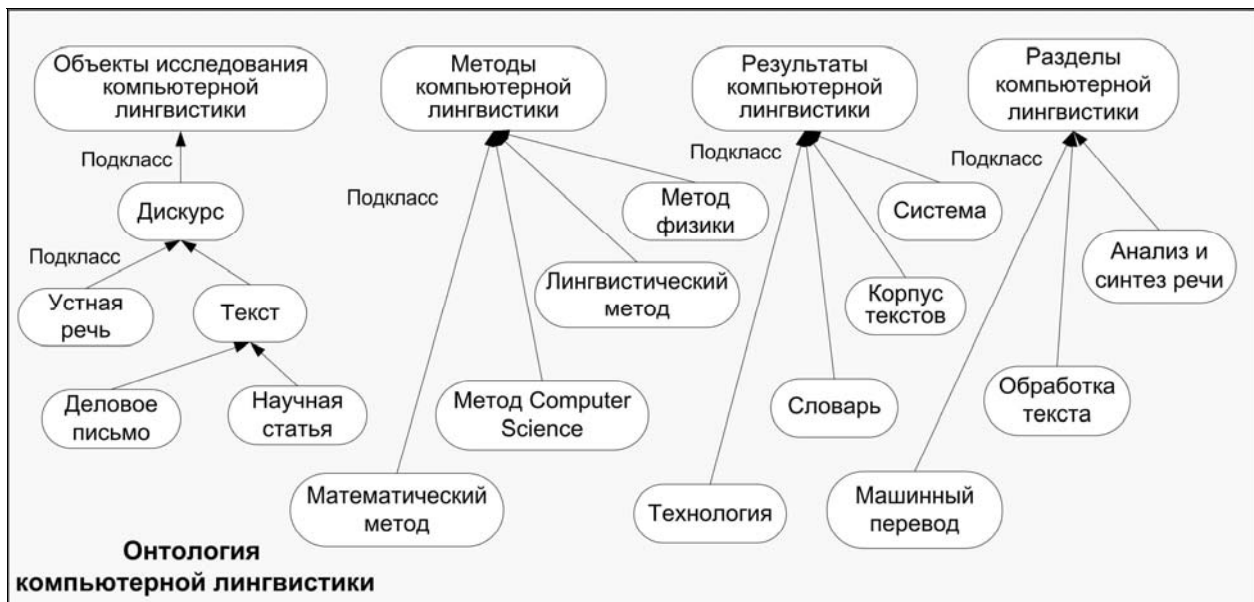


Рис.2. Ядро онтологии компьютерной лингвистики

Иерархия разделов науки определяет значимые разделы и подразделы компьютерной лингвистики. К разделам компьютерной лингвистики относятся, например, такие направления как Машинный перевод, Автоматическая обработка текста, Анализ и синтез речи и др. Эти общие направления также подразделяются на более частные. Например, Машинный перевод включает Автоматический и Автоматизированный машинный перевод.

Иерархия объектов исследования задает типизацию объектов исследования компьютерной лингвистики. В качестве базового объекта исследования рассматривается Дискурс как форма существования и использования языка (языковых единиц различных уровней в их системной взаимосвязи). В частности, рассматриваются такие формы Дискурса, как Текст и Устная речь.

Иерархия методов исследования служит для систематизированного описания инструментов исследования, применяемых в компьютерной лингвистике. Здесь выделяются такие подразделы, как Математические методы, Лингвистические методы, Методы Computer Science и др.

Иерархия научных результатов служит для типизации и описания результатов научной деятельности. Она включает такие типы результатов, как Технологии, Системы, Словари, Корпусы текстов.

Все иерархии онтологии компьютерной лингвистики, как и любой другой предметной онтологии, связаны между собой посредством ассоциативных отношений, семантика которых была определена при описании базовых онтологий.

3 Информационное содержание портала

Вводя формальные описания понятий предметной области в виде классов объектов и отношений между ними, онтология портала задает

структуры для представления реальных объектов и связей между ними. В соответствии с этим данные на портале представлены как множество разнотипных информационных объектов и связей, которые в совокупности образуют информационное содержание портала.

Информационный объект (ИО) – это структурированная совокупность данных, представляющая описание некоторого объекта выбранной области знаний или релевантного ей информационного ресурса. Каждый ИО соответствует некоторому классу онтологии (является экземпляром этого класса) и имеет заданную этим классом структуру.

Между конкретными информационными объектами могут существовать связи, семантика которых определяется отношениями, заданными между соответствующими классами онтологии.

Таким образом, информационное содержание портала (его контент) включает как знания общего характера (представлены в онтологии), так и конкретные знания о реальных объектах и информационных ресурсах (такие знания мы называем данными).

Описание информационных ресурсов является наиболее важным компонентом информационного наполнения портала. Как было сказано выше, каждый ресурс, представленный на портале, соответствует такому понятию онтологии, как *Информационный ресурс*. Набор атрибутов и связей ресурса основан на стандарте Dublin Core [21]. Его атрибутами являются: название, Интернет-ссылка (URL), язык, тип доступа и т.п. Описание ресурса хранится в БД портала и включает экземпляр данного понятия и набор экземпляров отношений, связывающих это понятие с другими понятиями онтологии (организациями, учеными, публикациями, событиями, разделами науки и т.д.).

4 Управление информационным контентом портала

Настройка портала на предметную область и управление его контентом осуществляется с помощью специализированных редакторов (редактора онтологии и редактора данных), реализованных как web-приложения и доступных зарегистрированным пользователям-экспертам через Internet.

Язык описания онтологий и редактор онтологий выбирались и проектировались таким образом, чтобы они были понятны экспертам-гуманитариям и ими было просто и удобно пользоваться. В частности, из-за этих требований мы отказались от использования таких популярных средств, как язык представления онтологий OWL [17] и редактор Protégé [18].

В качестве прообраза языка описания онтологий был взят уже хорошо зарекомендовавший себя в наших проектах язык представления знаний системы Semp-ТАО [5, 23]. Основной структурой для представления знаний и данных в этом языке является неоднородная семантическая сеть, состоящая из объектов, связанных направленными бинарными отношениями.

Объектом семантической сети может являться любая сущность предметной области, выделенная экспертом или инженером знаний. Каждый объект характеризуется своим именем и значениями атрибутов – слотов объекта. На слотах объекта могут быть заданы ограничения, представляющие собой логические выражения, связывающие значения слотов объекта. Объекты с одинаковыми свойствами объединяются в классы. На классах определено отношение наследования, образующее иерархию классов.

Особенностью отношений является то, что они могут иметь собственные атрибуты, специализирующие связь между аргументами, т.е. имеют вид:

$$R(Arg1, Arg2, Matr),$$

где R – имя отношения, $Arg1, Arg2$ – аргументы отношения (классы), $Matr$ – множество атрибутов, описывающих дополнительные свойства отношения.

Также отношениям могут быть приписаны такие математические свойства, как транзитивность, симметричность и рефлексивность.

Управление онтологиями осуществляется через редактор онтологий. С целью обеспечения распределенной разработки онтологий в этом редакторе поддерживается механизм делегирования прав экспертам разных уровней. С помощью редактора онтологий можно создавать, модифицировать и удалять любые элементы онтологии – классы, отношения, домены, а также задавать и модифицировать иерархии понятий.

Для более удобного представления информации пользователю портала в редактор онтологий включены средства настройки визуализации знаний

и данных. Эти средства позволяют для каждого класса онтологии задать шаблон визуализации объектов этого класса и шаблон визуализации ссылок на них.

Шаблон визуализации объектов класса определяет порядок, в котором отображаются все его атрибуты и связанные с ним отношения.

Для более наглядного и содержательного представления ссылки на конкретный объект класса шаблон ее визуализации может включать как атрибуты этого класса, так и атрибуты классов, связанных с ним отношениями, а также атрибуты этих отношений. Значения атрибутов, включенных в шаблон ссылки, используются для построения текстового представления ссылки на объект (гиперссылки).

Для того, чтобы можно было обмениваться онтологиями с другими информационными системами, а также для интеграции в портал знаний онтологий, разработанных другими исследователями, реализована и проходит отладку подсистема, выполняющая две функции: (1) конвертирование онтологии, представленной в формате портала знаний, в OWL-представление и (2) трансляцию онтологии, представленной в OWL-формате, во внутренний формат портала.

Ввод новых информационных объектов осуществляется с помощью управляемого онтологией редактора данных. Редактор данных позволяет создавать, редактировать и удалять информационные объекты и связи между ними. Формы для ввода конкретных ИО и их связей автоматически генерируются по онтологии портала.

Особенностью предложенного подхода является то, что портал знаний обеспечивает доступ не только к собственным информационным ресурсам, но и поддерживает навигацию по заранее размеченным (проиндексированным) ресурсам, размещенным в сети Интернет. При этом информация о ресурсах накапливается коллекционером онтологической информации о ресурсах, т.е. специальной подсистемой портала, осуществляющей сбор, анализ, оценку релевантности, автоматическое индексирование и классификацию близких тематике портала Интернет-ресурсов.

Коллекционер включает два модуля: модуль сбора информации (поисковый робот) и модуль индексирования и классификации ресурсов. Модуль сбора информации осуществляет поиск Интернет-документов по ссылкам, заданным экспертами и хранящимся в специальной базе данных, и определяет их релевантность тематике портала. Модуль индексирования и классификации, используя онтологию и предметный словарь, строит содержательный индекс для каждого документа и определяет раздел науки, к которому он относится (фактически, при этом создается информационный объект – экземпляр класса «Информационный ресурс» – со всеми его связями с другими информационными объектами).

ПРОСМОТР ОБЪЕКТА	
Проект	
Название деятельности	InDoc
Описание деятельности	Интеллектуализация документооборота: автоматическая классификация и аннотирование документов
Дата начала	2000
Дата окончания	2003
Стадия проекта	завершен
СВЯЗИ ОБЪЕКТА	
Исследует:	
Объект Исследования	
Деловое письмо	
Использует-Метод:	
Метод Исследования	
Модели представления знаний	
Семантический анализ	
Результат-Деятельности:	
Научный Результат	
Система InDoc	
Направление деятельности:	
Раздел Науки	
Автоматическая обработка текста	
Информационный поиск	
ССЫЛКИ НА ОБЪЕКТ	
Персона-Участник-Деятельности:	
Персона	Роль Участника Деятельности
Булгаков, С.В.	исполнитель
Загорюлько, Ю.А.	ответственный исполнитель
Кононенко, И.С.	исполнитель
Костов, Ю.В.	исполнитель
(Всего: 6)	
Организация-Участник-Деятельности:	
Организация	
РосНИИ искусственного интеллекта, РосНИИ ИИ	
Публикация о Деятельности:	
Публикация	
Кононенко, И.С., Сидорова, Е.А., Обработка делового письма в системе документооборота, 2002	
Загорюлько, Ю.А., Кононенко, И.С., Костов, Ю.В., Сидорова, Е.А., Подход к интеллектуализации документооборота, 2004	
Загорюлько, Ю.А., Кононенко, И.С., Костов, Ю.В., Сидорова, Е.А., Подход к разработке интеллектуальной системы документооборота инвестиционной компании, 2002	
Загорюлько, Ю.А., Кононенко, И.С., Костов, Ю.В., Сидорова, Е.А., Представление знаний в интеллектуальной системе документооборота, 2002	
(Всего: 6)	
Ресурс-Деятельности:	
Ресурс	
Проект InDOC	

Рис.3. Представление информационного объекта в виде HTML-страницы

5 Организация содержательного доступа на основе онтологий

Содержательный доступ к систематизированным знаниям и информационным ресурсам заданной области знаний обеспечивается с помощью предоставляемых порталом развитых средств навигации и поиска.

Основной сценарий работы пользователя с порталом состоит из выбора либо непосредственно с помощью средств визуализации, либо с помощью механизма поиска объектов определенного класса, их просмотра, навигации по их связям и фильтрации списков таких объектов.

5.1 Навигация

Для конечного пользователя данные на портале представлены в виде множества связанных информационных объектов. Вся информация о конкретном объекте и его связях отображается в виде HTML-страницы (Рис.3), формат и наполнение которой зависят от класса данного объекта и

заданного для него шаблона визуализации. При этом объекты, связанные с данным объектом, представляются на его странице в виде гиперссылок, по которым можно перейти к их детальному описанию.

Список объектов отображается в виде страницы, содержащей набор ссылок на эти объекты. Для больших списков формируется составная страница, включающая список страниц с элементами навигации по этому списку.

Навигация по данным портала представляет собой процесс перехода от одних информационных объектов к другим по заданным между ними связям.

Например, при просмотре информации о конкретном проекте (Рис.3) мы можем видеть значения его атрибутов и его связи с другими объектами. Используя представленные связи в качестве элементов навигации, можно перейти к просмотру подробной информации как по прямым связям (об описываемом объекте исследования, об используемых в проекте методах исследования и научных результатах), так и по обратным (об

участниках проекта, об информационном ресурсе, описывающем данный проект).

При переходе по конкретной связи любого информационного объекта мы можем получить достаточно большой список объектов (например, список всех участников крупного проекта или конференции). В связи с этим был введен механизм фильтрации списков информационных объектов. Фильтрация есть способ выборки подмножества ИО из списка путем наложения на него ограничений, т.е. задания фильтра.

Фильтр является набором условий, которые определяют допустимые значения атрибутов ИО и требования к существованию связей с другими информационными объектами. Этот метод позволяет, например, отфильтровать множество участников проекта как по возрасту или научной степени (условия на атрибут), так и по используемым ими методам исследования (условия на связанный объект).

5.2 Поиск

Поиск также базируется на онтологии, благодаря чему пользователю предоставляется возможность задания запроса в терминах предметной области портала. Основными элементами такого запроса являются понятия и отношения онтологии, а также ограничения, которым должны удовлетворять искомые данные.

Допустимые ограничения для атрибута зависят от типа его значений. Так, например, для атрибутов типа «число» (integer) и «дата» (data) может задаваться точное значение или допустимый интервал значений.

Для задания ограничений на объекты, связанные ассоциативными отношениями с искомым объектом, пользователю предоставляется возможность задать условия на значения всех атрибутов связанных объектов. При этом также могут быть заданы условия на значения атрибутов соответствующих отношений.

Например, запрос «Найти методы исследования, которые использовались для обработки деловых писем на русском языке в проектах в период с 1998 по 2005 год» будет выглядеть следующим образом:

Класс «Метод исследования»:

Отношение «Применяется к»:

Класс «Деловое письмо»

Атрибут «Язык» = «русский»

Отношение «Использует метод»:

Класс «Проект»

Атрибут «Дата начала»:

(>= 1998) & (<= 2005)

Атрибут «Дата окончания»:

(>= 1998) & (<= 2005)

На данный момент поисковые запросы задаются через специальный графический интерфейс, управляемый онтологией портала знаний. При выборе пользователем класса искомого информационных объектов автоматически генерируется поисковая форма, в которой можно

задать ограничения на значения атрибутов объектов выбранного класса, а также на значения атрибутов объектов, связанных с данным объектом ассоциативными отношениями.

6 Порталы знаний и семантические порталы

Использование онтологий и других элементов технологии Semantic Web при построении нашего портала знаний требует проведения его сравнения с так называемыми семантическими порталами (Semantic Web Portal), в настоящее время активно развиваемыми за рубежом [12]. В рамках этого направления исследований были созданы такие семантические порталы, как Esperonto [9], OntoWeb [19], MuseumFinland [11] и др.

Портал Esperonto [9] является примером практического применения технологии автоматической генерации порталов знаний ODESeW, разрабатываемой в Мадридском политехническом университете (Universidad Politecnica de Madrid). Он обслуживает информационные потребности участников европейского проекта Esperonto. Для данного портала средствами инструментальной системы WebODE были построены 5 различных онтологий: «Проект», «Документ», «Персона», «Организация» и «Встреча», которые служат для описания научно-исследовательских проектов и могут быть переиспользованы. Хотя эти онтологии и связаны несколькими отношениями, однако на портале Esperonto возможен поиск объектов только по атрибутам классов, представленных в рамках одной онтологии. Кроме того, данный портал ориентирован только на представление информации о научно-исследовательских проектах и не может быть настроен на другую область знаний.

Портал OntoWeb [16] относится к так называемым комьюнити порталам (community portal). Он также построен в рамках одноименного европейского проекта с целью облегчения передачи наработок по онтологиям и технологии Semantic Web от представителей научной общественности в индустрию. В онтологии портала представлены все необходимые для функционирования портала типы информационных объектов, поэтому участники проекта могут размещать на портале любую информацию об организациях, событиях, проектах, персонах и документах (научных статьях). Недостатком этого портала является то, что информация на нем организуется не по тематическому принципу, а исходя из текущих нужд проекта, из-за чего оказывается достаточно трудно найти необходимые документы.

Семантический портал “MuseumFinland – Finnish Museums on the Semantic Web” [14] обеспечивает интеграцию электронных коллекций основных музеев Финляндии. Декларируется, что он позволяет осуществлять навигацию и семантический поиск музейных экспонатов в этих

коллекциях на основе семи культурологических онтологий по девяти различным измерениям, таким как материал, тип артефакта, время создания, место использования, ситуация использования и т.д. Однако, наше исследование возможностей этого портала показывает ограниченность его средств поиска и навигации: пользователь может искать и использовать в качестве главного элемента навигации/поиска только сами музейные экспонаты. Т.е. данный портал не позволяет указать в качестве предмета поиска автора экспоната или материал, из которого он изготовлен, в то время как наш портал позволяет осуществлять поиск по любым понятиям онтологии и выполнять навигацию по любым ассоциативным отношениям.

Порталы, построенные в рамках предлагаемого нами подхода, имеют ряд особенностей, делающих их действительно порталами знаний. В частности, наши порталы знаний, в отличие от обычных семантических порталов, обеспечивают содержательный доступ не только к информационным ресурсам, относящимся к определенной области знаний, но и систематизированным знаниям как о самой области, так и об организации научной деятельности в ее рамках.

Заключение

В докладе представлен подход к организации содержательного доступа к гуманитарным информационным ресурсам на основе онтологий. Онтология предоставляет средства для эффективного представления разнообразной информации по заданной тематике, поддерживает систематизацию и интеграцию релевантных информационных ресурсов и содержательный доступ к ним.

По онтологии автоматически строятся:

- схема внутренней базы данных портала (логическая структура БД и ее ограничения целостности);
- формы для заполнения БД портала данными (информационными объектами, являющимися экземплярами понятий онтологии);
- схема навигации по информационному пространству портала (по отношениям онтологии);
- формы поисковых запросов (по понятиям и отношениям онтологии).

Благодаря использованию онтологий в качестве информационной модели, портал знаний представляет собой не просто еще один каталог ресурсов по заданной тематике, а сеть знаний и данных, по связям которой поддерживается удобная навигация и содержательный поиск.

Разделение онтологии портала на предметно-независимые и предметные онтологии делает портал настраиваемым на любую область научных знаний. А возможность декларативной подстройки

онтологии в ходе эксплуатации портала знаний позволит отслеживать динамику появления новых знаний и информационных ресурсов по тематике портала и, тем самым, обеспечит поддержку его актуальности и полезности.

На основе предложенного подхода совместно с Институтом археологии и этнографии СО РАН был разработан археологический портал знаний [1, 2], обеспечивающий содержательный доступ широкому кругу пользователей к систематизированным знаниям и информационным ресурсам по археологии и этнографии. Основой для построения предметной онтологии портала послужила предложенная в [7] системная классификация археологической науки, включающая более 500 понятий. В настоящее время археологический портал знаний содержит более 4 тысяч информационных объектов, связанных более чем 15 тысячами онтологических отношений, и продолжает активно пополняться новыми знаниями и данными.

Хотя первоначально данный подход разрабатывался для организации содержательного доступа к гуманитарным информационным ресурсам, по нашему убеждению, он может быть успешно применен и для других областей знаний. Возможно, при этом необходимо будет уточнить или дополнить базовые онтологии. В настоящее время исследуется возможность применения предложенного в докладе подхода для разработки Интернет-порталов знаний не только для гуманитарных, но и точных наук. В частности, этот подход сейчас применяется для разработки портала знаний по компьютерной лингвистике [4], которая представляет собой синтез гуманитарной (лингвистика) и точных (информатика и математика) наук.

Литература

- [1] О.А. Андреева, О.И. Боровикова, С.В. Булгаков, Ю.А. Загорюлько, Е.А. Сидорова, Б.Г. Циркин, Ю.П. Холюшкин. Археологический портал знаний: содержательный доступ к знаниям и информационным ресурсам по археологии // Труды 10-й национальной конференции по искусственному интеллекту с международным участием - КИИ'2006. –Москва: Физматлит, – 2006. – Т.3. – С.832-840.
- [2] Археологический портал знаний <http://www.sati.archaeology.nsc.ru/classarch2/>.
- [3] Боровикова О.И., Загорюлько Ю.А. Организация порталов знаний на основе онтологий // Компьютерная лингвистика и интеллектуальные технологии: Труды международного семинара «Диалог 2002» (Протвино, 6-11 июня 2002 г.). – Москва: Наука, – 2002. –Т.2, –С.76–82.
- [4] Загорюлько Ю.А., Боровикова О.И., Кононенко И.С., Сидорова Е.А. Подход к построению предметной онтологии для портала знаний по компьютерной лингвистике // Компьютерная

- лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2006» (Бекасово, 31 мая - 4 июня 2006 г.). – Москва: Изд-во РГГУ, 2006. С. 148-151.
- [5] Загорулко Ю.А., Попов И.Г.: Описание сложных предметных областей на основе интеграции средств представления знаний. Труды международного семинара Диалог'97 по компьютерной лингвистике и ее приложениям, Москва, 1997, с.110-115.
- [6] Смирнов С.В. Онтологии в прикладных интеллектуальных системах: прагматический подход // Труды 9-й национальной конференции по искусственному интеллекту с международным участием – КИИ–2004. – Москва: Физматлит, 2004. – Т.3. – С.1059-1067.
- [7] Холушкин Ю.П., Гражданников Е.Д. Системная классификация археологической науки (элементарное введение в археологическое науковедение). Новосибирск: Изд-во ИДМИ Минобразования. –Новосибирск. – 2000. –58 с.
- [8] Benjamins V. R., Fensel D., et. all, 1998, “Community is Knowledge! in KA2”, Proceedings of the KAW'98, Banff, Canada, 1998.
- [9] Esperanto
<http://www.esperanto.net/>
- [10] Corcho O., Gómez-Pérez A., López-Cima A., López-García V., MC. Suárez-Figueroa. ODESeW. Automatic Generation of Knowledge Portals for Intranets and Extranets. Lecture Notes in Computer Science Vol 2870. The Semantic Web - ISWC 2003. Springer-Verlag, October 2003, pp.802-817.
- [11] Hyvönen E., Salminen M., Kettula S., Junnila M. A Content Creation Process for the Semantic Web. Proceedings of OntoLex2004: Ontologies and Lexical Resources in Distributed Environments, May 29, 2004, Lisbon, Portugal.
- [12] Lausen H., Stollberg M., Hernández R. L., Ding Y., Han S.-K., and Fensel D. Semantic Web Portals - state of the art survey. Technical Report TR-2004-04-03, DERI (www.deri.org), 2004.
- [13] Mike Ushold, Michael Gruninger. Ontologies: Principles, Methods and Applications // Knowledge Engineering Review, Volume 11, Number 2, 1996.
- [14] MuseumFinland – Finnish Museums on the Semantic Web”
<http://museosuomi.cs.helsinki.fi>
- [15] Nicola Guarino. OntoSeek: Content-Based Access to the Web // IEEE Intelligent Systems, May/June, 1999, p. 70-80.
- [16] OntoWeb. Web site
<http://www.ontoweb.org/>
- [17] OWL Web Ontology Language Guide, 2004.
<http://www.w3.org/TR/owl-guide/>
- [18] Protégé. Web site
<http://protege.stanford.edu/>
- [19] Spyns P., Oberle D., Volz R., Zheng J., Jarrar M., Sure Y., Studer R., Meersman R.. OntoWeb a Semantic Web Community Portal. In Proc. Fourth International Conference on Practical Aspects of Knowledge Management (PAKM), December 2002, Vienna, Austria, LNAI 2569, Springer Verlag, pp.189-200.
- [20] Thomas R. Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing // Inter-national Workshop on Formal Ontology, March, Padova, Italy, 1993.
- [21] Using Dublin Core.
<http://dublincore.org/documents/usageguide/>
- [22] Zagorulko Yu., Borovikova O., Bulgakov S., Sidorova E. Ontology-based approach to development of adjustable knowledge internet portal for support of research activity // Bull. of NCC. Ser.: Comput. Sci. – 2005. – Is. 23.
- [23] Zagorulko Yury A., Popov Ivan G., Kostov Yury V.: Subdefinite Data Types and Constraints in Knowledge Representation Language. Joint Bulletin of the Novosibirsk Computing Center and Institute of Informatics Systems, Series: Computer Science, 16 (2001), NCC Publisher, Novosibirsk, 2001, p.153-170.

Ontology-Based Approach to Getting Content-Based Access to Humanitarian Information Resources

Zagorulko Yu., Borovikova O, Zagorulko G.

The paper presents approach to getting the content-based access to humanitarian information resources using ontology.

Ontology constitutes information basis of Internet knowledge portal that must provide both integration and systematization of humanitarian scientific knowledge and of information resources relevant to the subject domain of a portal and content-based access to them from any point of Internet space.

Ontology is used for automatic generation of scheme of internal data base of portal and forms for filling this data base, formulating user queries in terms of subject domain of portal and navigation through portal information space.

The structuring of portal ontology to domain-independent and subject domain ontologies, makes the knowledge portal easily adjustable to any area of knowledge.

* Данная работа выполняется при финансовой поддержке РГНФ (проект № 07-04-12149)