



OSTIS-2015

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

ОНТОЛОГИЧЕСКИЙ ПОДХОД К СОЗДАНИЮ НАУЧНЫХ ИНТЕРНЕТ-РЕСУРСОВ

Загорулько Ю.А., Загорулько Г.Б.

*Институт систем информатики им. А.П. Еришова Сибирского отделения
Российской академии наук, г. Новосибирск, Россия*

zagor@iis.nsk.su

gal@iis.nsk.su

В докладе представлен подход к построению научных интернет-ресурсов, предназначенных для информационной и аналитической поддержки научной и производственной деятельности в определенной области знаний. Особенностью подхода является использование онтологии для систематизации и интеграции знаний, информационных ресурсов, релевантных моделируемой области знаний, и средств интеллектуальной обработки содержащихся в них данных. Кроме того, на основе онтологии организуется удобная навигация по научным знаниям и информационным ресурсам, содержательный поиск данных и средств интеллектуальной обработки, а также строятся шаблоны для автоматического извлечения информации из Интернет.

Ключевые слова: научный интернет-ресурс, область знаний, информационный ресурс, онтология, специализированная программная оболочка.

Введение

Несмотря на то, что значительная часть информации практически по всем областям знаний представлена в Интернет, проблема эффективного обеспечения научного сообщества информацией по интересующим его тематикам еще далека до своего решения. Нерешенной остается и проблема доступа к методам и средствам обработки знаний и данных, собранных по этим тематикам. Большей частью, даже уже представленные в сети Интернет реализации методов обработки остаются недоступными из-за отсутствия содержательной информации о них, по которой их могли бы найти как программные агенты, так и члены научного сообщества.

Для представления знаний о некоторой области исследований и содержательного описания релевантных ей информационных ресурсов и методов обработки содержащихся в них данных и знаний удобным средством являются онтологии [Guarino, 19981] и построенные на их базисе семантические сети [Осипов, 1990; Загорулько, 2013]. Именно эти формализмы были положены в основу интеллектуальных систем, предназначенных для информационной и аналитической поддержки научной и производственной деятельности в определенных областях знаний. Такие системы

получили название тематические интеллектуальные научные интернет-ресурсы (ИНИР).

Создание тематических ИНИР позволит решить следующие задачи:

(1) сведение информационных ресурсов, относящихся к определенной области знаний, а также средств их интеллектуальной обработки в единое информационное пространство;

(2) поддержка логической целостности системы семантических описаний информационных ресурсов и средств их интеллектуальной обработки, сведенных в единое информационное пространство;

(3) обеспечение возможности открытого содержательного доступа не только к интегрированным информационным ресурсам, но и к средствам их интеллектуальной обработки.

Актуальность первой из этих задач обусловлена тем, что российская наука, образование и производство испытывают в наши дни потребность в концентрации и обобщении накопленной информации по различным отраслям знаний и эффективном ее использовании, но удовлетворение этой потребности затрудняется тем, что в силу своей многоплановости и многоаспектности научные информационные ресурсы рассредоточены на удаленных страницах множества сайтов, а также в распределенных электронных библиотеках и

архивах. Сведение в единое информационное пространство такого рода ресурсов, относящихся к требуемой области знаний, будет способствовать удовлетворению указанной потребности.

Поддержка логической целостности системы семантических описаний ресурсов и средств их интеллектуальной обработки необходима для обеспечения пользователей актуальной, достоверной и достаточно полной информацией по интересующей его теме. Решение этой задачи будет обеспечено тем, что в основе создаваемых ИНИР будут лежать онтологии, дающие целостное и непротиворечивое представление о моделируемой области знаний, а также семантические сети, построенные на концептуальном базисе онтологий.

Сведение информационных ресурсов и средств их интеллектуальной обработки в единое информационное пространство и обеспечение логической целостности системы их семантических описаний создаст предпосылки для решения третьей задачи – содержательного доступа к этим ресурсам и механизмам их интеллектуальной обработки широкому кругу пользователей.

Ввиду высокой потребности в системах такого класса нами разрабатывается технология создания и сопровождения ИНИР, ориентированная не на программистов и инженеров знаний, а непосредственно на специалистов в областях знаний, для которых такие ресурсы разрабатываются. Эта технология является развитием технологии [Загоруйко, 2011], успешно применявшейся при построении порталов научных знаний для многих предметных областей.

1. Концепция и архитектура тематического ИНИР

Тематический ИНИР представляет собой доступную через Интернет информационную систему, обеспечивающую систематизацию и интеграцию научных знаний и информационных ресурсов определенной области знаний, содержательный эффективный доступ к ним и средствам их интеллектуальной обработки. ИНИР позволяет исследователям значительно сократить время, требуемое для обеспечения доступа к необходимой информации и ее анализа, за счет аккумуляции описаний сущностей моделируемой области знаний и релевантных ей информационных ресурсов (в том числе, web-сервисов) непосредственно в контенте ИНИР.

Ядром системы знаний ИНИР является онтология, которая наряду с описанием моделируемой области знаний содержит соотнесенное с ним описание структуры и типологии интегрируемых информационных ресурсов и методов интеллектуальной обработки содержащихся в них данных.

Семантическая сеть, структура которой определяется онтологией ИНИР, играет роль

интеллектуального хранилища данных, в котором накапливается информация о релевантных научных информационных ресурсах и web-сервисах, реализующих методы обработки содержащихся в них знаний и данных.

На основе онтологии организуется удобная навигация по научным знаниям и информационным ресурсам, интегрированным в ИНИР, а также содержательный поиск данных и средств их интеллектуальной обработки.

Кроме онтологии и семантической сети в систему знаний ИНИР включен тезаурус, который рассматривается в качестве лексического дополнения онтологии. Он содержит описание терминов моделируемой области знаний, выраженных словами или словосочетаниями, с помощью которых понятия онтологии представляются в текстах и пользовательских запросах. Тезаурус задает смысл понятий посредством соотнесения одних терминов с другими, главным образом, с помощью семантических отношений. Благодаря этому он может применяться как при обработке пользовательских запросов, так и при поиске, обработке и аннотировании информационных ресурсов, интегрируемых в ИНИР.

ИНИР имеет традиционную для информационных систем трехуровневую архитектуру (см. Рис. 1), включающую уровень представления информации, уровень обработки информации и уровень хранения и доступа к информации (базовый уровень).

Первый уровень обеспечивается пользовательским интерфейсом, главными функциями которого являются представление пользовательских запросов и результатов поиска и решений задач, а также обеспечение управляемой онтологией навигации в информационном пространстве ИНИР. При этом пользовательский интерфейс обеспечивает не только содержательный доступ к контенту ИНИР, но и к средствам аналитической обработки информации.



Рисунок 1 – Архитектура тематического ИНИР

На уровне обработки информации обеспечиваются различные виды поиска и обработки информации, а также ее передача между уровнями. Для этих целей данный уровень включает средства поиска информации в контенте ИНИР, а также средства ее аналитической обработки, реализованные, в том числе, и в виде web-сервисов [McIlraith, 2001].

Средства поиска позволяют осуществлять как поиск информации по ключевым словам, так и расширенный семантический поиск с представлением запроса в терминах понятий и отношений онтологии и ограничений на них. Эти же средства поддерживают навигацию по контенту ИНИР, поставляя в пользовательский интерфейс семантическую окрестность просматриваемых понятий и объектов.

В качестве аналитических инструментов используются средства фильтрации и визуализации понятий онтологии и объектов семантической сети. Фильтрация позволяет из большого списка выбрать объекты, значения атрибутов которых удовлетворяют указанным ограничениям. Сервис визуализации [Апанович, 2011] обеспечивает отображение полной системы понятий онтологии в виде графа, предоставляет пользователю возможность оценить объем каждого понятия с помощью круговой диаграммы (т.е. количество объектов, относящихся к данному понятию), позволяет графически отобразить либо полную сеть объектов, либо только граф объектов, связанных выбранным отношением.

Базовый уровень обеспечивает выполнение функций хранения и управления знаниями (онтологией и тезаурусом) и данными (контентом ИНИР) с использованием средств реляционных СУБД (MySQL), технологий Semantic Web (OWL, RDF) [OWL, 2004] и семантических web-сервисов (WSDL, OWL-S) [Chinnici, 2007; Martin et al., 2004].

2. Методологическая и техническая поддержка построения ИНИР

В настоящее время разрабатывается технология построения ИНИР. Как было сказано выше, ее особенностью является ориентация на широкий круг пользователей – экспертов, т.е. специалистов в определенных областях знаний. Такая технология позволяет экспертам собирать и систематизировать в рамках единого информационного пространства обширные знания и данные, релевантные требуемой области знаний, а также средства их интеллектуальной обработки.

Так как основу ИНИР составляют онтологии, центральным элементом технологии их создания является методология построения онтологий, к которой примыкает методология построения тезаурусов. Технология базируется на программных средствах, поддерживающих построение онтологий, тезаурусов и управление контентом ИНИР.

2.1. Методология построения онтологий ИНИР

Онтология конкретного ИНИР строится в соответствии с методологией, главными принципами которой являются:

- структурирование онтологии ИНИР путем разделения ее на относительно независимые онтологии, каждая из которых представляет самостоятельный компонент знаний;
- использование базовых онтологий, которые включают только самые базовые сущности, не зависящие от области знаний ИНИР;
- построение всех онтологий ИНИР на основе базовых онтологий путем их доработки и развития.

Использование такой методологии значительно упрощает создание онтологии ИНИР и ее дальнейшее сопровождение.

Как было сказано выше, онтология ИНИР кроме описания понятий и отношений моделируемой области знаний включает соотношенное с ним описание структуры и типологии интегрируемых информационных ресурсов и методов интеллектуальной обработки данных и знаний. В связи с этим онтология ИНИР состоит из трех онтологий, отвечающих за представление указанных выше компонентов знаний, а именно: онтологии области знаний ИНИР, онтологии научных информационных ресурсов и онтологии задач и методов.

В качестве базовых онтологий предложены онтология научной деятельности и онтология научного знания, на основе которых строится онтология области знаний ИНИР, а также базовые онтологии научных информационных ресурсов и задач и методов.

Онтология научного знания содержит классы, задающие структуры для описания понятий конкретных областей знаний, такие как Раздел науки, Метод исследования, Объект исследования, Научный результат и др. Эта онтология также включает отношения, связывающие между собой объекты указанных выше классов.

Онтология научной деятельности включает классы понятий, относящиеся к организации научной и исследовательской деятельности, такие как Персона, Организация, Событие, Научная деятельность, Проект, Публикация и др. Она включает также отношения, связывающие понятия данной онтологии как между собой, так и с понятиями онтологии научного знания.

Базовая онтология научных информационных ресурсов включает класс Информационный ресурс в качестве основного класса. Этот класс служит для описания, релевантных области знаний информационных ресурсов. Набор атрибутов и связей этого класса основан на стандарте Dublin core [Hillmann, 2005]. Его атрибутами являются: название ресурса, язык ресурса, тематика ресурса,

тип доступа к ресурсу и т.п. Объекты этого класса могут быть связаны семантическими отношениями с другими информационными объектами, представляющими в контенте ИНИР организации, персоны, публикации, события, разделы науки и т.д.

Базовая онтология задач и методов включает такие классы как Задача, Метод и Web-сервис, а также отношения, связывающие эти понятия между собой и понятиями других базовых онтологий. С помощью понятий и отношений данной онтологии могут быть описаны задачи, для решения которых предназначен ИНИР, методы их решения, а также реализующие их web-сервисы.

Описания web-сервисов онтологии задач и методов базируются на онтологии OWL-S [Martin et al., 2004], предназначенной для описания семантических web-сервисов. Благодаря этому с web-сервисом связываются не только описание его интерфейса в терминах типов входных и выходных данных, но и описание его семантики, т.е. того, что сервис делает, его предметной области, ограничений на область применения и качество сервиса и т.п. Наличие семантического описания у web-сервисов обеспечивает не только реализацию их поиска и корректного использования (исполнения), но и создает предпосылки и для их успешной интеграции в ИНИР. При этом будет обеспечиваться содержательный доступ к ним не только для программных агентов, но и для исследователя, желающего найти необходимые для решения его задач средства интеллектуальной обработки информации.

2.2. Построение тезауруса

Методология построения тезауруса области знаний ИНИР базируется на тех же принципах, что и методология построения онтологии.

Так, тезаурус области знаний ИНИР строится путем дополнения ядра тезауруса, содержащего описания терминов, представляющих понятия базовых онтологий, в том числе в интернет-ресурсах.

Ядро тезауруса, как и весь тезаурус, строится на основе онтологии представления тезауруса [Загоруйко, 2012], которая включает набор базовых понятий и отношений, присутствующих в любом тезаурусе. В частности, она содержит классы, описывающие такие сущности тезауруса, как термины, которые подразделяются на дескрипторы (предпочтительные термины) и аскрипторы (текстовые входы, которые при поиске и индексировании документов могут быть заменены на соответствующие дескрипторы), источники терминов (web-ресурсы, текстовые документы или коллекции текстов, в которых встречаются или определяются термины) и области/подобласти знаний, с которыми могут быть соотнесены термины. В онтологии также представлены отношения, связывающие объекты перечисленных выше классов между собой.

2.3. Управление системой знаний ИНИР

Поддержка процесса управления системой знаний ИНИР обеспечивается редакторами онтологий и данных, а также подсистемой сбора онтологической информации из Интернет.

Указанные редакторы реализованы как web-приложения, поэтому поддерживают удаленную коллективную разработку и поддержку системы знаний ИНИР авторизованными экспертами через Интернет.

Редактор онтологий предназначен для построения онтологий и управления ими в процессе жизненного цикла ИНИР. Этот редактор проектировался таким образом, чтобы им могли пользоваться эксперты, не являющиеся специалистами в области информатики и программирования.

Управление контентом ИНИР осуществляется с помощью редактора данных. Этот редактор работает под управлением онтологии, что позволяет не только значительно облегчить ввод данных (путем генерации удобных форм для ввода объектов и связей между ними), но и обеспечить их логическую целостность.

Построение тезауруса и последующее редактирование его содержания осуществляется с помощью того же редактора данных, но работающего уже под управлением онтологии представления тезауруса, что позволяет обеспечить логическую целостность его терминологической системы.

Для того чтобы ИНИР был полезным ресурсом его система знаний должна содержать достаточно полную информацию о моделируемой области знаний и выполняемой в рамках ее научной и/или производственной деятельности. Создание такого ИНИР – довольно трудоемкая задача, требующая значительных усилий разработчиков. Для ее автоматизации предназначена подсистема сбора информации из Интернет, реализующая методы поиска и извлечения информации, базирующиеся на онтологии и механизмах метапоиска.

Сбор информации для ИНИР включает следующие задачи: поиск релевантных его области знаний интернет-ресурсов и документов, извлечение информации из них и занесение полученной информации в контент ИНИР. Для выполнения этих задач в подсистему сбора информации входят модуль поиска релевантных интернет-ресурсов, модуль извлечения информации из интернет-ресурсов, модуль занесения найденной информации в контент ИНИР, а также база данных ссылок на интернет-ресурсы (БД СИР).

При настройке ИНИР на область знаний эксперт заполняет БД СИР ссылками на релевантные по его мнению интернет-ресурсы. При этом для каждой ссылки указывается класс онтологии, объекты которого описывает соответствующий ей ресурс.

Список ссылок может пополняться не только вручную – экспертами, но и автоматически – модулем поиска интернет-ресурсов, который выполняет сбор ссылок на релевантные интернет-ресурсы по поисковым запросам, сформированным на основе названий классов онтологии и терминов тезауруса, представляющих понятия моделируемой области знаний. Этот модуль запускается с заданной при настройке ИНИР периодичностью. При этом модуль поиска обращается к поисковым системам Google, Яндекс и Bing через их программные интерфейсы, т.е. использует механизм метапоиска с последующей фильтрацией дубликатов и нерелевантных ссылок [Ахмадеева, 2013].

Контент ИНИР заполняется информацией из таких источников, как порталы знаний, сайты организаций, ассоциаций, проектов и конференций, новостные ленты, социальные научные сети, вики-ресурсы, реестры (каталоги) веб-сервисов и др. Из этих источников извлекается информация о проектах, организациях, персонах, конференциях и публикациях, т.е. обо всех объектах классов онтологии научной деятельности, а также о самих источниках, как объектах класса Информационный ресурс онтологии научных информационных ресурсов.

Для каждого из этих классов разрабатывается свой метод извлечения информации, включающий набор шаблонов и связанных с ними обработчиков. В шаблонах для каждого типа извлекаемой информации указываются обработчики, реализующие алгоритмы обхода и/или анализа соответствующих фрагментов интернет-страниц или документов. Указанные шаблоны генерируются на основе онтологии, а затем уже при необходимости дополняются экспертами. Для повышения полноты извлечения информации увеличивается вариативность этих шаблонов за счет использования в них альтернативных терминов из тезауруса (синонимов и гипонимов).

Например, шаблон для извлечения информации о проектах, построенный на основе описания класса Проект, имеет вид, представленный на рисунке 2.

```
<Class Name="Проект" engine="FragmentSearch">
  <Marker Term="О проекте" PType="Menu/Head" FragType="Page/Block"/>
  <Marker Term="Проект" PType="Head" FragType="Block"/>
  <Attr Name="Название" type="string" engine="NameEntity">
    <Marker Term="Проект" PType="link" FragType="LinkText"/>
    <Marker Term="Проект" PType="sentence" FragType="QuoteText"/>
    <Marker Term="Проект" PType="Head" FragType="Head"/>
  </Attr>
  <Attr Name="Аннотация" type="text">
    <Marker Term="Аннотация/Содержание проекта/Описание
    проекта/ О проекте" PType="Head" FragType="Block/Page"/>
  </Attr>
  <Relation Name="Публикация о Проекте">
    <Marker Term="Публикации" PType="Menu/Head" FragType="Page/Block"/>
    <Marker Term="Список публикаций" PType="Menu" FragType="Page"/>
    <Marker Term="Литература" PType="Menu/Head" FragType="Page/Block"/>
    <Marker Term="Библиография" PType="Menu/Head" FragType="Page/Block"/>
    <Object Name="Публикация" engine="PublicationsList"/>
  </Relation>
  <Relation Name="Участник проекта">
    <Marker Term="Об участниках" PType="Menu" FragType="Page"/>
    <Marker Term="Список участников" PType="Head" FragType="Block"/>
    <Marker Term="Исполнители" PType="Head" FragType="Block"/>
    <Marker Term="Участники" PType="Head" FragType="Block"/>
    <Object Name="Персона" engine="PersonList"/>
  </Relation>
</Class>
```

Рисунок 2 – Шаблон для извлечения информации о проекте

Шаблон содержит указание на то, для извлечения объектов какого класса он предназначен (блок Class). Он также включает блоки атрибутов (Attr), отношений (Relation) и аргументов отношений (Object). Каждый из этих блоков может описываться одним или группой альтернативных маркеров (Marker), задающих свойства фрагмента текста, содержащего извлекаемую информацию.

Заключение

В докладе представлен подход к построению тематических ИНИР, обеспечивающих систематизацию и интеграцию информационных ресурсов определенной области знаний и средств интеллектуальной обработки содержащейся в них информации, а также содержательный эффективный доступ к ним и их использование при решении различных задач.

Важным преимуществом ИНИР является то, что он позволяет исследователям значительно сократить время, требуемое для доступа к необходимой информации и ее анализа, за счет аккумуляции описаний сущностей моделируемой области знаний и релевантных ей информационных ресурсов непосредственно в своем контенте.

Использование онтологии в качестве основы ИНИР, создает предпосылки для того, чтобы технология построения ИНИР стала действительно массовой. С одной стороны, онтология является удобным средством формирования и фиксации общего разделяемого экспертами-разработчиками знания о данной предметной области, обеспечивая при этом возможность переиспользования знаний, что упрощает и ускоряет разработку новых приложений. С другой стороны, базирование средств описания области знаний ИНИР, как и создания и сопровождения его контента, на онтологии делает их доступными для использования непосредственно экспертами, так как представление знаний и данных в виде объектов и отношений между ними, принятое в онтологии, является наиболее естественным для человека.

Чтобы данной технологией мог воспользоваться широкий круг экспертов, в ее рамках должны быть разработаны программные оболочки разных типов ИНИР, отличающиеся набором базовых онтологий и, возможно, программных компонентов. Такие специализированные оболочки будут представлять собой «пустые» ИНИР, т.е. в них будут представлены все необходимые структурные компоненты будущего ИНИР, но не достроены нижние уровни онтологии области знаний и не заполнен контент.

Примером такой специализированной оболочки является оболочка портала научных знаний [Загорулько, 2008], включающая рассмотренные выше базовые онтологии – онтологию научной деятельности и онтологию научного знания, с помощью которых эксперт может, не прибегая к помощи инженеров знаний и программистов,

построить онтологию требуемой области знаний. С помощью этой оболочки были построены научные интернет-порталы по археологии [Андреева, 2006] и компьютерной лингвистике [Боровикова, 2006].

Работа выполнена при финансовой поддержке РФФИ проект № 13-07-00422 «Методы и технологии создания и управления интеллектуальными научными Интернет-ресурсами на основе онтологий и семантических сетей».

Библиографический список

[Андреева, 2006] Андреева О.А., Боровикова О.И., Булгаков С.В. и др. Археологический портал знаний: содержательный доступ к знаниям и информационным ресурсам по археологии // Тр. 10-й национальной конференции по искусственному интеллекту с международным участием КИИ-2006 (25–28 сентября 2006 г., Обнинск). М.: Физматлит, 2006. Т. 3. С. 832–840.

[Апанович, 2011] Апанович З.В., Винокуров П.С., Кислицина Т.А. Методы и средства визуализации информационного наполнения больших научных порталов // Вестник НГУ Серия: Информационные технологии. 2011. — Т.9, выпуск 3. — С. 5-14.

[Ахмадеева, 2013] Ахмадеева И.Р., Загорулко Ю.А., Саломатина Н.В., Серый А.С., Сидорова Е.А., Шестаков В.К. Подход к формированию тематических коллекций текстов на основе интернет-ресурсов // Вестник НГУ. Серия: Информационные технологии. 2013. Т.11, выпуск 4.

[Боровикова, 2006] Боровикова О.И., Загорулко Ю.А., Загорулко Г.Б. и др. Разработка портала знаний по компьютерной лингвистике // Тр. 11-й национальной конференции по искусственному интеллекту с международным участием КИИ-2008 (Дубна, 2008 г.). М.: ЛЕНАНД, 2008. Т. 3. С. 380–388.

[Загорулко, 2008] Загорулко Ю.А., Боровикова О.И. Подход к построению порталов научных знаний // Автометрия. № 1, 2008, т. 44. — с. 100–110.

[Загорулко, 2011] Загорулко Ю.А. Технология построения порталов научных знаний: опыт применения, проблемы и перспективы // Труды 21-й Международной Крымской конференции «СВЧ-техника и телекоммуникационные технологии» – КрыМиКо-2011 –Севастополь, Крым, Украина, изд. Севастополь: Вебер, 2011. –Т.1. –С.51–54.

[Загорулко, 2012] Загорулко Ю.А., Боровикова О.И. Программная оболочка для построения многоязычных тезаурусов предметных областей, ориентированная на экспертов // Труды 13-й национальной конференции по искусственному интеллекту с международным участием КИИ-2012. – Белгород: Изд-во БГТУ, 2012. -Т.4. -С. 76-83.

[Загорулко, 2013] Загорулко Ю.А. Технологии разработки интеллектуальных систем, основанные на интегрированной модели представления знаний // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2013): материалы III Междуна. научн.-техн. конф. (Минск, 21-23 февраля 2013 г.) / редкол.: В. В. Голенков (отв. ред.). – Минск: БГУИР, 2013. –С. 31-42.

[Осипов, 1990] Осипов Г.С. Построение моделей предметных областей. Неоднородные семантические сети // Известия АН СССР. Техническая кибернетика. –1990. – №5. – с. 32–45.

[Chinnici, 2007] Chinnici R., Moreau J.-J., Ryman A., Weerawarana S. Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language. W3C Recommendation 26 June 2007. Available at: <http://www.w3.org/TR/wsd20/> (accessed 1 December 2014).

[Guarino, 1998] Guarino N. Formal Ontology in Information Systems // Proceedings of FOIS'98 (Trento, Italy, 1998). Amsterdam: IOS Press, 1998. pp. 3-15.

[Hillmann, 2005] Hillmann D. Using Dublin Core. 2005. Available at: <http://dublincore.org/documents/usageguide/> (accessed 8 May 2014).

[Martin et al., 2004] Martin D. et al. OWL-S: Semantic Markup for Web Services. W3C Member Submission 22 November 2004.

Available at: <http://www.w3.org/Submission/OWL-S/> (accessed 1 December 2014).

[McIlraith, 2001] McIlraith S.A., Son T.C., Zeng H. Semantic Web Services. Intelligent Systems, IEEE, 2001, 16(2). pp. 46-53.

[OWL, 2004] OWL Web Ontology Language Guide. W3C Recommendation 10 February 2004. Available at: <http://www.w3.org/TR/owl-guide/> (accessed 1 December 2014).

ONTOLOGICAL APPROACH TO DEVELOPMENT OF SCIENTIFIC INTERNET RESOURCE

Zagorulko Yu.A. *, Zagorulko G.B. *

** A.P. Ershov Institute of Informatics Systems
Siberian Branch of the Russian Academy of
Sciences, Novosibirsk, Russia*

zagor@iis.nsk.su

gal@iis.nsk.su

The paper presents an approach to development of intelligent scientific Internet resources (ISIR) intended to providing information and analytical support of research and production activity in certain area of knowledge. An important advantage of ISIR is its ability to appreciably reduce time of access to information required by researchers and processing it due to accumulation of descriptions of entities of the modelled knowledge area and Internet resources relevant it directly in the ISIR content.

A main feature of the approach is using ontology for systematization and integration of knowledge, information resources and intelligent processing facilities relating to the modeled area of knowledge. Besides, a convenient navigation through scientific knowledge and information resource. content-based access to data and intelligent processing facilities as well as information extraction from Internet are also organized on the basis of ontology.