

УДК 020.23.25

Е.В. Тимофеева, А.Ю. Шевченко

## МЕТОДЫ И ТЕХНОЛОГИИ ИСПОЛЬЗОВАНИЯ WEB-ОНТОЛОГИЙ В ОБРАЗОВАТЕЛЬНОМ ПРОЦЕССЕ

### 1. Введение

В течение последних нескольких лет необычайное влияние интернет-технологий на все сферы применения вычислительной техники породило новую тенденцию, заключающуюся в том, что традиционные методы представления информации отошли на второй план по сравнению с методами представления информации в форме, удобной для ее использования в среде Интернет.

Целью данной статьи является формирование базовых принципов построения онтобазируемых приложений для образовательного процесса.

При этом задачей статьи является исследование возможностей использования технологии OWL для построения онтобазируемых приложений, решающих задачи образовательного процесса.

В ходе развития образовательной системы и перехода к Болонской конвенции, данное направление заняло приоритетные позиции. На текущих этапах развития нашего государства наблюдается стремительная интеграция в Европейский Союз, что влечет значительные изменения в таких аспектах, как образовательная система государства. Следствием этого будут все новые и новые изменения в документообороте с целью преобразования в европейский стандарт. Чтобы успешно справиться с задачей подведения документооборота под новые стандарты и требования следует прибегнуть к онтобазируемому подходу, который предоставит возможность создания базовой онтологии. Применение онтонаправленных систем обеспечит качественно новый уровень поиска и интеграции информационных источников, сделает возможным создание интеллектуальных систем, способных формировать новые знания и данные.

В последнее время достаточно активно развивается технология Интранет, основная идея которой — построение информационной системы организации на базе web-протоколов и, таким образом, включение этой системы в Интернет. Все разработчики коммерческих СУБД уже включили и активно развивают поддержку импорта и экспорта данных с использованием технологии XML. Более того, достаточно обоснованно и аргументированно обсуждается использование XML, как основного формата хранения данных. Если будет сделан соответствующий выбор в пользу XML, то это приведет к очень серьезным изменениям.

Соответственно это новое веяние затронуло и сферу технологий искусственного интеллекта. Раз-

витие технологии RDF на базе XML, позволило значительно упростить взаимодействие интересов специалистов и объединить специфические формы и форматы представлений объектов и процессов в таких предметных областях как представление знаний, обработка естественно-языковой информации, принятие решений, экспертные системы и др. на базе форматов интернет-технологий. Как видно на практике, взаимодействие интернет-технологий практически с любой научной областью значительно ускоряет темпы ее развития, что подтверждается, в частности, очевидным в данный момент интересом к проблеме представления знаний. Одной из самых последних разработок в этой области, основанной на направлении Semantic Web, поддерживаемом www-консорциумом, является язык web-онтологий OWL.

По причине того, что знания являются неотъемлемой частью любого образовательного процесса, возникает достаточно закономерный и естественный вопрос о том, как можно наилучшим образом использовать результаты этих исследований в интернет-технологиях обучения, когда предполагается весьма косвенное взаимодействие учителя и обучаемого вследствие множества форм общения, предоставляемого интернет-технологиями. В связи с этим приводится краткий обзор языка OWL и затрагивается вопрос о возможностях его использования для информационной поддержки образовательного процесса в ближайшем будущем.

Таким образом, актуальность данной задачи определяется тем, что в настоящее время существует необходимость перехода от систем со статистической обработкой информации к системам с возможностью семантического анализа.

### 2. Технологии языка web-онтологий (OWL)

В рамках курса философии известно, что термин «онтология» используется для обозначения системы знаний, которые относятся к окружающему нас миру. Иначе говоря, онтология — это наука о бытии, наука о природе вещей и взаимосвязях между ними. Если же рассматривать термин «онтология» в рамках информационных технологий представления знаний, то данным термином можно определить определенный механизм, способ, который используется для описания некоторой области знаний или же предметной области. В приложениях к web-технологиям, можно отметить, что основной целью онтологий является представле-

ние (отображение) смысла понятий, используемых в конструкциях RDF. Соответствующим образом, RDF — это язык общения программных систем, работающих в среде Интернет, а онтологии, таким образом, составляют его словарь-тезаурус. Для достижения этой цели язык web-онтологий должен решить задачи следующего характера:

— во-первых, предоставить такой синтаксис описания понятий, который был бы достаточно удобен всем участникам взаимодействия в сети Интернет с учетом программных систем и человеческого фактора;

— во-вторых, сделать как можно более выразительными методы описания рассматриваемых понятий и связей между ними;

— в-третьих, обеспечить механизмы предполагаемого эволюционирования описаний и совместного использования онтологий в среде Интернет.

Теперь, наиболее уместным будет рассмотрение того, как решается каждая из этих трех задач в рамках языка OWL.

Учитывая то, что язык OWL является составной частью системы технологий Semantic Web, он, так же как и RDF, использует в качестве своего синтаксиса язык XML. Это позволяет решить первую задачу, т. к. теговые конструкции, которые активно и достаточно эффективно используются в высокоуровневых web-протоколах (языках HTML, XML), являются определенным образом оптимальными соглашениями в настоящее время для форм представления информации, которые ориентируются одновременно и на программные системы, и на человека. Использование стандартного для Интернет синтаксиса позволяет использовать уже существующие xml-анализаторы для синтаксического разбора web-онтологий и сконцентрировать внимание, тем самым, на анализе их семантической структуры. Учитывая тот факт, что синтаксис языка OWL является довольно стандартным по своей структуре, а также с учетом его ориентированности на web, он может уравнивать недостатки web-онтологий и явиться решающим при его выборе в качестве механизма представления знаний конкретной предметной области. Тем более, что современная наука и образование, так же как и другие сферы человеческой деятельности, подвергаются влиянию всеобщих значительных глобальных изменений. Стандарт www-консорциума на язык XML уже переведен на русский язык и доступен для ознакомления [1].

Решение второй поставленной нами задачи непосредственно связано с методами и средствами представления знаний. На данный момент существует несколько подходов к описанию объектов и явлений окружающего мира. Соответственно, появилось и достаточно большее количество языков представления знаний, реализующих эти подходы.

В языке web-онтологий используется в несколько упрощенном виде один из наиболее популярных сегодня и понятных на интуитивном уровне подходов — объектно-ориентированный метод описания окружающего мира. Основные идеи языка OWL — это его свойства, классы, объекты и ограничения, которые реализуют представление о мире, как о множестве объектов, которые характеризуются некоторым определенным набором свойств. Эти объекты состоят между собой в определенных отношениях и объединяются по определенным признакам (характерным свойствам и ограничениям) в группы (классы).

Таким образом, при создании онтологии мы начинаем с описания иерархий классов понятий, составляющих данную предметную область. Представлен небольшой фрагмент такого описания для онтологии «Высшие учебные заведения»:

```
<owl:Class rdf:ID=«Высшее Учебное Заведение»/>
<owl:Class rdf:ID=«Университет»>
<rdfs:subClassOf rdf:resource=«# Высшее Учебное
Заведение»/>
</owl:Class>
<owl:Class rdf:ID=«Харьковский Университет
Радиоэлектроники»>
<rdfs:subClassOf rdf:resource=«#Университет»/>
</owl:Class>
```

Для того, чтобы понятия предметной области несли определенную смысловую нагрузку, они должны характеризоваться конкретными наборами свойств и состоять в определенных связях друг с другом. Эту задачу в языке OWL решают механизмы свойств и ограничений, которые с ними ассоциируются. Свойства, в свою очередь, делятся на два вида: свойства-характеристики и свойства-связи. Первые характеризуют объекты (классы) и принимают в качестве своих значений данные определенных типов. Вторые связывают объекты (классы) друг с другом и соответственно принимают в качестве своих значений объекты (классы). На свойства накладываются ограничения двух типов: локальные и глобальные. К глобальным ограничениям относятся домены (классы, объекты которых могут обладать этими свойствами) и диапазоны (классы, объекты которых могут выступать в качестве значений этих свойств). Локальные ограничения накладываются на свойства в рамках определенного рассматриваемого в конкретном случае класса и могут еще больше сужать диапазоны для свойств в рамках этого класса, определять мощность свойств и их виды.

После того, как мы описали все классы, свойства, ограничения и объекты предметной области, мы получаем сложную систему иерархий, являющуюся своего рода базой для построения программ-

ных систем, которые обладают способностью осуществлять операции определенного интеллектуального уровня над информацией, содержащейся в онтологии. К таким операциям можно отнести, например, семантический поиск или определение целостности и достоверности информации на основе ограничений, которые закладываются в онтологию.

Если рассматривать язык OWL с точки зрения полноты выразительных возможностей, то можно отметить, что он не в состоянии конкурировать со многими существующими языками представления знаний уже хотя бы по той причине, что в текущей версии не рассчитан на описание стремительно изменяющихся явлений, учитывая то, что его основа – это объекты и связи. В то же время, тот факт, что этот язык является, в свою очередь, элементом стека веб-протоколов, т. е. является актуальным в рамках развития информационных технологий, позволяет рассчитывать, что в ближайшее время ограничения на его выразительные возможности будут сняты.

Однако есть одна задача веб-онтологий, связанная с поддержкой их развития и совместного использования, которая решается при помощи введения специальных механизмов в язык OWL. К ним, в частности, относятся механизмы описания версий онтологий и механизмы обработки данных, содержащихся в онтологиях. Первые позволяют описать ссылки на предыдущие версии онтологий и изменения, сделанные в данной версии по отношению к предыдущей. Это позволяет отслеживать и, таким образом, фиксировать эволюцию онтологий, а также работать с онтологиями разных авторов для одной и той же предметной области. Механизмы агрегирования данных решают задачи объединения различных онтологий, размещенных в сети Интернет, в частности, задачи ссылок из одной онтологии на классы и объекты другой онтологии. Все это позволяет выстраивать онтологии в единую базу знаний, в которую в ближайшем будущем и должен превратиться Интернет. В этом случае, программные системы будут способны разобраться и сориентироваться в глобальной сети и осуществлять различные интеллектуальные операции над размещенной в ней информацией.

Таким является наше краткое, но достаточно обзорное изложение языка описания веб-онтологий OWL. Наш обзор ставил своей целью скорее заинтересовать новой технологией и описать ее возможности в рамках ближайшего ее развития. Соответственно, для дальнейшего ознакомления есть возможность обратиться к материалам, размещенным на сайте www-консорциума, где можно найти достаточно большой объем информации относительно рассмотренной технологии.

Производя оценку достоинств и недостатков языка OWL, следует еще раз отметить тот факт, что этот язык является развитием технологий XML и RDF и входит в разрабатываемый сегодня новый стек веб-протоколов. Наряду с замечательными возможностями по совместному использованию OWL-онтологий, этот факт, на наш взгляд, может стать решающим при выборе языка OWL в качестве механизма построения базы знаний или экспертной системы. Это преимущество позволяет не акцентировать внимание на невысокие пока выразительные возможности этого языка, особенно, учитывая скорости появления новых версий программных интернет-приложений. Следует также отметить, что в отличие от многих языков представления знаний, язык OWL не является языком, представляющим, в основном, теоретический интерес и мало пригодным для решения практических задач. В его развитие вкладываются немалые средства, для его поддержки создается множество программных инструментов и систем. И уже сегодня есть примеры его практического использования. Поэтому стоит всерьез рассмотреть перспективы и возможности построения базы знаний своей дисциплины, курса, науки, потому как инструмент для этого уже существует. И использовать как представленную в базе информацию, так и ее структуру как инструменты технологий обучения. О том, как можно использовать подобную базу знаний представляется нижеизложенная информация.

### 3. Возможности использования языка web-онтологий в рамках образовательного процесса

На данный момент существует множество определений понятия база знаний, но следует остановиться на следующем, удовлетворяющем нас, определении: база знаний – это база данных плюс мета-данные. Под мета-данными в данном конкретном случае понимаются правила синтаксической и семантической обработки данных. Таким образом, база знаний является как источником информации, так и, весомой базой для построения программных систем, включая программных агентов, способных обрабатывать эту информацию. Исходя именно из этого, и нужно искать направления использования базы знаний в конкретной предметной области (в нашем случае – в технологиях обучения). Известно, что в рамках многих учебных процессов на определенном уровне существуют проблемы несогласованности дисциплин. Так, при системном изучении компьютерных технологий система курсов делится на аппаратный цикл, цикл операционных систем и системного программного обеспечения, цикл прикладного программного обеспечения, цикл вычислительных сетей. На сегодняшний день, пожалуй, наиболее отработанным,

с точки зрения своей структуры, является аппаратный цикл. Остальные же циклы имеют чрезвычайно сложную и запутанную структуру. Особенно это касается цикла прикладного программного обеспечения и вычислительных сетей с учетом появления постоянно совершенствующихся новых технологий. Между тем, все эти дисциплины, курсы и циклы связаны множеством прямых и обратных связей, и должны составлять единую, понятную, хорошо структурированную систему. Понимание этой системы — необходимое условие для современного специалиста в области информационных технологий. В этой связи возникает вопрос — почему бы не попробовать создать некую базу знаний, которая станет закрепляющим средством, механизмом, объединяющим огромное количество существующей информации в единую систему, основой для ориентирования в области компьютерных технологий. Конечно, построить такую сложную систему одному человеку или одной кафедре вряд ли удастся. Однако механизмы языка OWL позволяют строить базу знаний, начиная со своей дисциплины, со своего цикла. Подобная база знаний может стать для преподавателей — прежде всего, ориентиром при построении новых и развитии старых курсов, а для студентов — принципиально новым, с точки зрения своих возможностей, источником информации. В далекой перспективе, возможно, будут построены программные агенты, которые смогут автоматически пополнять подобную базу знаний информацией, вновь появляющейся в Интернет, таким образом, автоматически поддерживая ее актуальность. Таково, на наш взгляд, основное направления использования языка OWL в учебном процессе. Второе направление использования языка OWL связано с построением программных систем, которые оперируют знаниями, содержащимися в онтологиях. При наличии базы знаний дисциплины или курса, мы уже сегодня можем строить тестирующие программные системы, которые по своим возможностям, очевидно, превзойдут существующие на сегодняшний день тесты, ориентированные на выбор одного из нескольких вариантов. Новые тестирующие агенты будут способны сами сформулировать вопрос и оценить ответ на него с точки зрения ограничений, описанных в онтологии. Эти агенты могут быть, очевидно, использованы так же и для тренировки умений и навыков, и, в свою очередь, для учебного моделирования. Эти два направления использования web-онтологий в учебном процессе, на наш взгляд, являются сегодня наиболее очевидными, но ими не ограничивается список возможностей онтологий. Как инструмент для описания знаний, онтологии должны заинтересовать всех специалистов, сталкивающихся в своей практике с проблемой представления и использования знаний. К этим специалистам, вне всякого сомнения, относятся и преподаватели.

#### 4. Язык OWL в рамках семантической сети

Сегодня на первом плане разработок в сфере Internet-стандартов находится Семантическая Сеть (Semantic Web, согласно терминологии W3C), архитектура которой предполагает наличие у любой информации, находящейся в сети, связанный с этой информацией точный смысл, который нельзя было бы перепутать даже в случае совпадения фраз или слов, встреченных в разных контекстах. Фактически это означает, что любая информация связана с некоторым неотделимым от нее контекстом. Семантическая Сеть активно использует язык XML для определения собственной структуры документов и язык RDF (Resource Definition Framework), предоставляющий удобную среду формализации метаданных и сведений о контексте. RDF создан консорциумом W3C и предназначен для описания метаданных, является подмножеством языка XML и имеет собственный язык RDF Schema для описания структуры документов. Однако RDF — это самый низкоуровневый из существующих языков описания метаданных, поскольку оперирует лишь понятиями связей примитивных сущностей. Со временем разработчикам Семантической Сети стало очевидно, что средств XML и RDF для представления информации и метаданных для построения полноценной семантически связанной сети недостаточно. RDF подобен ассемблеру, если сравнивать семантическую нагрузку отдельных конструкций языка, и слишком сложен для решения задачи подобного масштаба. Чистый XML, в свою очередь, являясь метаязыком, включает в себя RDF как подмножество, и не создан для какого-либо конкретного применения, а потому для построения семантической сети его также недостаточно. Поэтому консорциумом W3C и был создан язык онтологий OWL (Web Ontology Language) [5].

Онтология определяет термины, с помощью которых можно описать предметную область. Использование онтологий особенно необходимо в приложениях-агентах, осуществляющих поиск и объединение информации из различных источников и из разных сред, в которых один и тот же термин может означать разные вещи. Несмотря на то, что DTD (Document Type Definition, формальное описание структуры XML-документов) в стандарте XML и схем XML (XML Schema) вполне достаточно для обмена данными между сторонами, которые заранее договорились о значении определений и терминов, отсутствие семантики в указанных средствах описания структуры серьезно ограничивает надежность выполнения задачи поиска и объединения данных при использовании новых XML-словарей. Например, элемент <Rate>, встретившийся индексактору в разных документах, может означать либо курс валюты, либо оценку за экзамен, либо цену товара, или что-нибудь еще. И без точной

информации о том, что именно в конкретном документе имеется в виду под этим элементом, поисковый агент не сможет со стопроцентной вероятностью вернуть именно то, что и требовалось пользователю.

Практически любой пользователь Сети хотя бы раз сталкивался с ситуацией, когда при поиске интересующей его информации он помимо прочего получал от поисковой машины множество бесполезных ссылок. Поскольку поиск информации осуществляется вне контекста, никакие уточнения запросов не смогут надежно найти именно то, что нужно. Для качественного осуществления поиска пользователю необходимо понимать все тонкости предметной области, включая ее лексику, термины, определения, иерархии сущностей — одним словом, досконально знать онтологию. Хорошо, если пользователь является экспертом предметной области, хотя даже в этом случае будет найдена лишняя информация. Для обычного пользователя язык OWL призван упростить процесс поиска, возложив необходимость знания предметной области и описание контекста поиска полностью на авторов документа и систему поиска, причем передача этих функций авторам документа должна быть незаметна для пользователя.

### 5. Заключение

В статье описаны принципы построения онтологии образовательного процесса на языке OWL, а также предложены методы построения онтобазированных систем для образовательного процесса.

Новизной, фактически, является применение новых семантических методов для обработки разнородной слабоструктурированной информации, относящейся к образовательному процессу.

Новая технология предоставляет такие возможности для обработки, описания и менеджмента информационных ресурсов. При использовании этой технологии одним из главных и необходимых компонентов является единая общая онтология пред-

метной области, в данном случае — образовательного процесса.

Благодаря применению онтобазированных систем появляется возможность семантической обработки информации. Этот подход позволяет легко интегрировать информацию и формировать на основе разработанных онтологий общую информационно-онтологическую базу. Это позволит создать распределенные интеллектуальные приложения для поддержки систем, работающих в различных областях, в том числе и в образовательном процессе.

В данный момент имеет место заметный переход к интеллектуальным системам поиска, обработки и анализа информации, основным механизмом которого является онтобазированный подход. Применение онтонаправленных систем обеспечит качественно новый уровень поиска и интеграции информационных источников, сделает возможным создание интеллектуальных систем, способных формировать новые знания и данные.

Таким образом, в рамках данной технологии, появилась возможность поиска, анализа и обработки информации, связанной с организацией учебного процесса на профилирующей кафедре, которая отображает иерархию заданных сущностей и связей между ними в машинопонятном, структурно стандартизированном виде.

**Список литературы:** 1. [Тим Брэй и др., 2000] Тим Брэй, Джин Паоли, С.М. Сперберг-Макквин, Ив Мэйлер. 2. Расширяемый язык разметки (XML) 1.0 (вторая редакция). Перевод Радика Усманова. // <http://www.citforum.ru/internet/xmlspec/index.shtml>. [Michael K. Smith et al., 2003] Michael K. Smith, Chris Welty, Deborah McGuinness. 3. OWL Web Ontology Language Guide // <http://www.w3.org/TR/owl-guide/>. 4. Semantic Web Activity. <http://www.w3.org/2001/sw>. 5. OWL Web Ontology Language 1.0 Reference. W3C Working Draft <http://www.w3.org/tr/owlref/default.htm>. 6. Проект: «Protégé 2000» <http://protege.stanford.edu>. 7. Д. Люгер. Искусственный интеллект: стратегии и методы решения сложных проблем, 4-е издание: Пер. с англ. — М.: Издательский дом «Вильямс», 2003. — 864 с.

*Поступила в редколлегию 16.11.2005*