Содержание

Посвящается Константине, Вангелису, Гиоргосу
– Григорис Антониоу

Посвящается профессору Томасу Гроту
– Паул Грос

Предисловие к серии

Традиционный взгляд на информационные системы как на дорогостоящие, сделанные на заказ программные приложения, основанные на базах данных, быстро меняется. Это изменение обусловлено, с одной стороны, развитием индустрии программного обеспечения, в которой все чаще используются готовые типовые компоненты и стандартные программные решения, а с другой – натиском информационной революции. В свою очередь, эти изменения привели к формированию новых требований к информационным услугам: однородность с точки зрения представления и взаимодействия их моделей, открытость с точки зрения архитектуры программного обеспечения и глобальность с точки зрения масштаба использования. Эти требования пришли в основном из прикладных областей, таких как электронная коммерция, банковское дело, производство (в том числе собственно индустрия программного обеспечения), профессиональное обучение, образование и управление природоохранной деятельностью; и здесь перечислены лишь немногие из этих областей.

Информационные системы будущего должны поддерживать бесперебойное взаимодействие с большим количеством независимых друг от друга источников данных различных производителей и унаследованных приложений, работающих на разнородных платформах в распределенных информационных сетях. При описании содержимого таких источников данных и при их интеграции решающую роль будут играть метаданные.

Кроме того, информационные системы следующего поколения должны поддерживать больше разнообразных социально ориентированных моделей взаимодействия. Эти модели могут включать модели навигации по данным, запросов к данным, извлечения данных и должны работать в сочетании с механизмами персонализированных уведомлений, аннотирования и профилирования. Кроме того, такие модели взаимодействия должны эффективно поддерживаться прикладными программными продуктами и быть динамически интегрированы в единые совместно используемые среды, ориентированные на пользователей. Кроме того, нужны крупные инвестиции как правительств, так и бизнеса в информационные ресурсы, подразумеваются конкретные меры по обеспечению безопасности, соблюдению принципа неприкосновенности частной жизни и правильности контента этих ресурсов.

Все перечисленные задачи должны быть решены в информационных системах следующего поколения. Мы называем такие системы *совместными информационными системами* (англ. *cooperative information systems*), и именно они находятся в центре внимания этой серии.

Говоря простым языком, совместные информационные системы удовлетворяют разнообразным сочетаниям требований, определяемых *контентом – обществом – коммерцией*. Эти требования порождаются текущими тенденциями в области готовых программных решений, таких как ERP-системы и системы электронной коммерции.

Основной задачей при создании совместных информационных систем является разработка технологий, что предполагает непрерывное совершенствование и увеличение массовых вложений сил и средств в информационные ресурсы и системы. Такие технологии должны обеспечивать соответствующую инфраструктуру, которая поддерживает не только разработку, но и развитие программного обеспечения.

Результаты ранних исследований в области совместных информационных систем становятся основой технологий, используемых для разработки социально ориентированных информационных порталов или шлюзов. Информационный портал обеспечивает реализацию принципа «одного окна» для доступа к разнообразным информационным ресурсам и услугам, тем самым создавая сообщество постоянных пользователей.

Научно-исследовательские результаты, которые послужат базой для реализации совместных информационных систем, не могут быть получены в рамках только одного научного направления в области информационных технологий. Представляется перспективным использование таких достаточно зрелых технологий, как базы данных, системы, основанные на знаниях, распределенные системы, программное обеспечение поддержки коллективной работы и графические пользовательские интерфейсы. Несмотря на то что отдельные технологии еще требуют усовершенствования, предполагается, что основным фактором успеха в этой области будет эволюция существующих технологий в единую технологию для разработки и управления совместными информационными системами.

Серия MIT Press «Совместные информационные системы» включает в себя учебники и научные издания, посвященные данной области и предназначенные для исследователей и специалистов, которые хотят быть в курсе текущих событий и будущих тенденций.

Серия будет включать в себя три типа книг:

- учебники и справочные пособия, предназначенные для студентов старших курсов и аспирантов;
- научно-исследовательские монографии, в которых собраны и обобщены результаты исследований и опыт разработок последних лет;
- сборники научных трудов, включающие статьи на конкретные темы.

Мы ждем предложений от авторов. Для этого надо предоставить в редакцию оглавление предлагаемой книги и образец глав книги. Все предоставленные предложения будут формально рассмотрены, и авторы будут уведомлены о результатах этого рассмотрения.

Данные в источнике данных являются полезными, так как они моделируют некоторую часть реального мира, объект изучения (или приложение, или область интерпретации). Проблема *семантики данных* заключается в установлении и поддержании соответствия между источником данных, который в дальнейшем будем называть моделью, и предполагаемым объектом моделирования. Моделью могут являться база данных о сотрудниках в компании, схема базы данных, описывающей детали, проекты и поставщиков, веб-сайт, представляющий информацию об университете, или текстовый файл, описывающий битву при Ватерлоо. Проблема семантики данных появилась с момента разработки первых баз данных. Однако эта проблема может оставаться под контролем, пока базы данных функционируют в закрытой и относительно стабильной среде. В этих условиях смысл данных определяется правильно определенной структурой базы данных, и предполагается, что он будет интерпретироваться небольшим количеством постоянных пользователей и прикладных программ.

Появление веба все изменило. Базы данных сегодня стали в некоторой форме доступными в вебе, который является открытой средой: пользователи, прикладные программы, способы использования данных постоянно меняются. В этих условиях семантика данных должна быть доступна непосредственно вместе с самими данными. Доступность семантики для пользователей можно обеспечить за счет выбора соответствующего формата представления информации. Однако для прикладных программ семантика должна быть представлена в формальной, машинно обрабатываемой форме. Именно этот факт обусловил появление инициативы под названием «семантический веб»[1].

---

[1] Tim Berners-Lee and Mark Fischetti, Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor. San Francisco: HarperCollins, 1999.

Неудивительно, что это инициатива Тима Бернерса-Ли привлекала огромное внимание как исследователей, так и практиков. В настоящее время регулярно проводится международная конференция, посвященная семантическому вебу «International Semantic Web Conference»[1], издательством «Elsevier» выпускается журнал «Journal of Web Semantics»[2], действуют отраслевые комитеты, которые рассматривают стандарты первого поколения для семантического веба[3].

Данная книга представляет собой своевременную публикацию, учитывая стремительный характер развития концепций, технологий и стандартов семантического веба. Книга предлагает постепенное введение в концепцию и технологии семантического веба, в том числе XML, DTD, XML-схемы, RDF, RDFS, OWL, логику и логический вывод. Все материалы книги сопровождаются примерами и приложениями, иллюстрирующими способы использования этих технологий.

Мы рады включить эту книгу, посвященную семантическому вебу, в серию «Совместные информационные системы». Надеемся, что для читателей она будет интересной, познавательной и полезной.

Джон Милопоулос
(John Mylopoulos)
jm@cs.toronto.edu
Кафедра компьютерных наук
Университет Торонто
Торонто, Онтарио
Канада

Майкл Папазоглоу
(Michael Papazoglou)
M. P. Papazoglou@kub.nl
INFOLAB
P. O. Box 90153
Университет Тилбурга
Нидерланды

---

[1] http://swsa.semanticweb.org/content/international-semantic-web-conference-iswc.

[2] www.semanticwebjournal.org.

[3] На момент перевода книги уже утвержден целый ряд стандартов семантического веба второго поколения. См. http://www.w3.org/standards/semanticweb/. – *Прим. перев.*

Глава 1

Концепция семантического веба

1.1. Введение

1.1.1. Цели сематического веба

Общая концепция «семантического веба» может быть сведена к одной фразе: «сделать веб более доступным для компьютеров». На данный момент веб – это веб текста и изображений. Такие мультимедийные материалы полезны для людей, но компьютеры играют очень ограниченную роль в вебе: они индексируют сайты по ключевым словам, передают информацию от серверов к клиентам, но этим и ограничиваются их функции. Вся интеллектуальная работа (выбор информации, ее объединение, агрегирование и т. д.) выполняется человеком. А что, если сделать веб более информативным для машин, наполнить его машиночитаемыми, то есть «понятными» для машин, данными? Такой веб позволит реализовать функции, осуществление которых невозможно на данный момент. *Поиск* информации не будет больше ограничиваться просто поиском по ключевым словам, он может стать более осмысленным, учитывать синонимы, исключать омонимы, а также учитывать контекст и цель поисковых запросов. Сайты могут стать более *персонализированными*, если браузеры смогут понимать содержимое веб-страницы и адаптировать его к личным интересам пользователя в соответствии с его профилем. *Связывание* страниц может стать более осмысленным, если реализовать динамическое создание полезных ссылок на основе действий конкретного пользователя, вместо того заранее создать одинаковые ссылки для всех пользователей. Станет возможна *интеграция* информации с разных веб-сайтов, в отличие от настоящего времени, когда пользователю для объединения найденной на сайтах информации приходится использовать «интеллектуальную» технологию копипаста.

1.1.2. Проектные решения сематического веба

Существуют различные походы к построению более «семантического» веба. Один из них – разработка проекта «Giga Google», основанного на «чрезмерно высокой эффективности данных»[1] для поиска правильных взаимосвязей между словами, между терминами и контекстом и т. д. Отсутствие прогресса в производительности поисковых машин, свидетелями чего мы являемся в течение нескольких последних лет, позволяет предположить, что такой поход имеет ограничения: ни один из поисковых гигантов не смог предложить решение, позволяющее в результате поиска возвращать нечто большее, чем простой список несвязанных страниц.

Семантический веб, также известный в последние годы как веб данных (англ. *Web of Data*)[2], основан на различных принципах проектирования, которые можно обобщить следующим образом:

1) обеспечение доступности в вебе структурированных и полуструктурированных данных, представленных в стандартных форматах;
2) обеспечение доступности в вебе не только наборов данных, но и отдельных элементов данных и отношений между ними;
3) описание предполагаемой семантики таких данных с помощью формализма, обеспечивающего возможность машинной обработки этой семантики.

Решение использовать структурированные и полуструктурированные данные основано на результатах исследований, а именно на том факте, что на сегодняшний день основой неструктурированного «веба текста и изображений» является на самом деле очень большое количество структурированных и полуструктурированных документов.

В основном содержание современных веб-страниц генерируется на основе баз данных и систем управления контентом, содержащих хорошо структурированные наборы данных. Однако структурные связи, существующие в таких наборах данных, полностью теряются при публикации их в виде человекочитабельных веб-страниц на языке разметки гипертекста HTML (Hypertext Markup Language) (см.

---

[1] The Unreasonable Effectiveness of Data Alon Halevy, Peter Norvig, and Fernando Pereira, IEEE Intelligent Systems, March/April 2009, pgs 8-12, http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en//pubs/archive/35179.pdf.

[2] http://www.readwriteweb.com/archives/web_of_data_machine_accessible_information. php.

рис. 1.1). Для того чтобы сделать веб более «семантическим», необходимо реализовать следующую ключевую идею: публиковать и связывать друг с другом структурированные наборы данных (вместо того чтобы публиковать и связывать между собой HTML-страницы, после того как потеряна большая часть структурных связей данных).

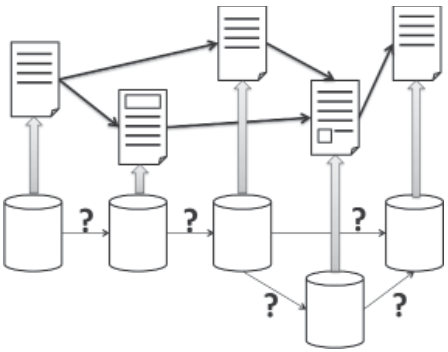

Рис. 1.1 ❖ Структурированные и неструктурированные данные в вебе

1.1.3. Основные технологии семантического веба

Три вышеупомянутых принципа проектирования реализованы с помощью конкретных технологий, описанию которых и посвящена большая часть данной книги:

1. В качестве модели данных для описания объектов и отношений между ними используются *помеченные графы*. Объекты представляются как вершины графа, а отношения между ними – как дуги. В качестве формализма для представления таких графов используется язык с не очень удачным названием – фреймворк описания ресурсов «Resource Description Framework» RDF[1].
2. Для идентификации отдельных элементов данных и отношений между ними, которые включаются в наборы данных, используются *веб-идентификаторы* URI (Uniform Resource Identifier)[2].

[1] Возможно, более удачным названием для языка было бы название «Rich Data Format» (многофункциональный формат данных).

[2] В последних версиях стандартов сематического веба для идентификации ресурсов используется IRI, а не URI. IRI (Internationalized Resource Identifier) – интернационализированный идентификатор ресурса, представляет собой Unicode-строку и соответствует синтаксису, определённому в стандарте URI. – *Прим. перев.*

Использование веб-идентификаторов отражено в стандарте языка RDF.

3. В качестве модели данных, позволяющей формально представить предполагаемую семантику этих данных, используются *онтологии* (если говорить кратко: иерархически организованные словари терминов – типов объектов и отношений между ними – свойств). Данная модель представляется с помощью таких формализмов, как язык RDF Schema и язык веб-онтологий OWL (Web Ontology Language), в которых URI используются для идентификации терминов и их свойств.

1.1.4. От данных к знаниям

Важно понимать, что такие формализмы, как RDF Schema и OWL, являются не только языками описания данных, но и фактически простыми языками *представления знаний*, поскольку они обеспечивают представление предполагаемой семантики. Эти языки являются «логикой», которая позволяет на основе явно заданной информации получать (выводить) дополнительную информацию. Язык RDF Schema представляет собой логику с очень низкой выразительностью, которая позволяет получать только очень простые выводы, такие как наследование свойств в соответствии с иерархией терминов и ограничениями на область значения и область определения свойств. Язык OWL представляет собой более выразительную (но все еще относительно простую) логику, которая дает возможность осуществлять более сложные выводы, например о равенстве и неравенстве, определенных ограничениях, о существовании объектов и т. д. Возможность осуществлять подобные выводы в языках RDF Schema и OWL позволяет разработчикам сайтов публиковать минимально возможное количество фактов, которые читатели должны знать. Кроме того, язык OWL дает возможность описать публикуемую информацию таким образом, чтобы читатель не мог предполагать (додумывать) определенные факты об опубликованных данных (имеется в виду – не нарушая согласованности опубликованной онтологии).

Таким образом, возможность реализации логического вывода для этих языков позволяет задавать как нижнюю, так и верхнюю границу предполагаемой семантики публикуемых данных. По мере уточнения онтологии эти нижние и верхние границы могут сближаться, тем самым все более точно обозначая предполагаемую семантику. Близость такого сближения определяется только точностью онтологии, требуемой для конкретного варианта ее использования.

1.1.5. Веб-архитектура сематического веба

Одна из основных характеристик традиционного веба заключается в том, что веб-контент является распределенным как в смысле расположения, так и в смысле владения им: веб-страницы, которые ссылаются друг на друга, часто размещены на разных веб-серверах, и эти серверы находятся в разных местах и принадлежат разным лицам. Развитие веба основано на принципе «кто угодно может сказать что угодно о чем угодно»[1], или точнее: любой может сослаться на любую веб-страницу без какого-либо разрешения владельца этой страницы и не заботясь о правильности ее адреса. Аналогичный механизм работает и в сематическом вебе (см. рис. 1.2): одна сторона может опубликовать набор данных в вебе (левая часть диаграммы), вторая сторона – независимо опубликовать словарь терминов (правая часть диаграммы), а третья сторона может проаннотировать данные первой стороны с помощью терминов из словаря второй стороны, не спрашивая разрешения ни у одной из сторон. Более того, первые две стороны могут даже не знать об этом. Именно подобное ослабление связей является сутью паутинообразной природы сематического веба.

Рис. 1.2 ❖ Веб-архитектура связанных данных

1.1.6. Из настоящего в будущее

Для того чтобы реализовать вышеизложенную концепцию семантического веба и соответствующие архитектурные решения, требуется осуществление следующих существенных шагов:

1) необходимо договориться о стандартном синтаксисе для представления данных и метаданных;
2) нужно иметь соглашение о словарях метаданных, для того чтобы совместно использовать предполагаемую семантику данных;

[1] http://www.w3.org/DesignIssues/RDFnot.html.

3) следует опубликовать большие объемы данных в форматах, предложенных на шаге 1, с использованием словарей, разработанных на шаге 2.

За последние десять лет (самые ранние проекты сематического веба датируются последними годами XX века) в реализации всех трех шагов был достигнут существенный прогресс: языки RDF, RDF Schema и OWL (и их вариации, такие как RDFa, OWL2 и т. д.) получили официальную поддержку консорциума World Wide Web Consortium (W3C) и приобрели статус стандартов де-факто. В этих форматах опубликованы тысячи словарей[1], и уже начинает происходить сближение между ним за счет развития автоматизированных технологий отображения онтологий и под давлением социальных и экономических требований (например, разработки словаря schema.org)[2]. Рост облака связанных данных LOD (Linked Open Data))[3] привел к тому, что миллиарды объектов и отношений между ними становятся доступны в вебе на основе общего синтаксиса и словарей.

1.1.7. Где мы сейчас

По сравнению с ситуацией на момент публикации первого издания этой книги в 2003 году, многие из основных элементов семантического веба уже реализованы. Разработаны быстро развивающиеся технологии для поддержки всех этапов развертывания приложений сематического веба, стремительно растет количество значительных проектов семантического веба как для коммерческих, так и для общественных организаций. Однако основные проблемы пока остаются неразрешенными: всевозрастающие объемы данных, преодоление барьеров, возникающих при внедрении, и, конечно, борьба с вездесущим врагом информационных систем: семантической гетерогенностью (неоднородностью).

1.2. Технологии сематического веба

1.2.1. Явное описание метаданных

В настоящее время веб-контент форматируется в расчете на то, что он будет анализироваться человеком, а не программами. Большинство веб-станиц разработано на языке HTML (непосредственно или

---

[1] http://swoogle.umbc.edu/.
[2] http://schema.org.
[3] http:/linkeddata.org.

с помощью соответствующих инструментов). Часть типичной веб-страницы, содержащей информацию о физиотерапевтическом центре, может быть представлена следующим образом:

```
<h1> Физиотерапевтический центр "Гибкость" </ h1>
Добро пожаловать на главную страницу физиотерапевтического центра "Гибкость".
Вы чувствуете боль? У Вас были травмы? Наши сотрудники Лиза Дэвенпорт, Келли
Таунсенд (наш очаровательный секретарь) и Стив Мэтьюз позаботятся о Вашем
теле и душе.
<h2> Приемные часы </ h2>
Пн. 11.00 - 19.00 <br>
Вт. 11.00 - 19.00 <br>
Ср. 15.00 - 19.00 <br>
Чт. 11.00 - 19.00 <br>
Пт. 11.00 - 13.00 <p>
Но обратите внимание, что прием не ведется во время проведения ежегодного со-
ревнования
<a href = …> State of Origin </a>.
```

Для людей эта информация представлена в удобочитаемом виде, но при ее анализе машинами могут возникнуть проблемы. Поиск на основе ключевых слов идентифицирует слова «физиотерапевтический» и «Приемные часы». Интеллектуальный агент может даже идентифицировать персонал центра. Но при этом могут возникнуть проблемы с тем, чтобы отличить врачей от секретаря, и еще большие проблемы – с определением точного времени приема (для решения которых необходимо перейти по соответствующей ссылке, чтобы найти информацию о днях проведения соревнования State of Origin).

Подход к решению этих проблем в рамках концепции семантического веба не подразумевает разработку суперинтеллектуальных агентов. Вместо этого предлагается решать данные проблемы со стороны веб-страниц. Веб-страницы смогут представлять структуру и смысл их контента, если заменить язык HTML на языки, более подходящие для этих целей. В этом случае, кроме информации о форматировании, предназначенной для отображения страниц, пригодных для чтения людьми, страницы могли бы содержать информацию об их контенте.

Первым шагом в этом направлении является создание расширяемого языка разметки XML (eXtensible Markup Language), который позволяет определять структуру информации веб-страниц. В нашем примере информация на языке XML может быть представлена следующим образом:

```
<company>
  <treatmentOffered>Физиотерапия</treatmentOffered>
```

```
  <companyName> Физиотерапевтический центр "Гибкость" </companyName>
  <staff>
    <therapist> Лиза Дэвенпорт </therapist>
    <therapist>Стив Мэтьюз</therapist>
    <secretary>Келли Таунсенд</secretary>
  </staff>
</company>
```

Это представление более пригодно для машинной обработки. В частности, это может быть использовано для обмена информацией в вебе, который является одной из самых известных областей применения технологии XML.

Однако язык XML по-прежнему представляет уровень синтаксиса, так как он позволяет описать структуру информации, а не ее смысл. Основным языком сематического веба является язык RDF, который позволяет представить информацию в виде утверждений. Для нашего примера подобные утверждения могут быть записаны следующим образом:

```
Компания А предлагает физиотерапевтическое лечение.
Название компании А "Физиотерапевтический центр Гибкость".
Лиза Давенпорт врач.
Лиза Дэвенпорт работает в компании А.
…
```

Для человека, просматривающего веб-страницу, разница между представлением информации на языке XML и списком RDF-утверждений может показаться минимальной, но эти два формата отличаются по своей природе: XML позволяет описать структуру данных, а RDF позволяет формулировать утверждения от этих данных[1].

Термин «*метаданные*» означает «данные о данных». Метаданные позволяют частично представить значение, смысл данных, а следовательно, и слово «сематический» в словосочетании «семантический веб».

1.2.2. Онтологии

Термин «онтология» происходит из философии. В этом контексте он означает название одной из областей философского знания, а именно «учение о бытии» или «учение о сущем» (буквальный перевод греческого слова Οντολογία). Учение о бытии – это раздел метафизики, изучающий наиболее общие категории сущего и способы их описа-

[1] Человек, читающий XML-страницу, интерпретирует значение ее элементов на основе имен тегов; очевидно, что машина не может этого сделать.

ния. Приведем пример типичного онтологического суждения: мир состоит из определенных объектов, которые могут быть сгруппированы в абстрактные классы на основании их общих свойств.

Однако в последние годы термин «онтология», как и многие другие слова из различных областей, стал использоваться в информатике и приобрел техническое значение, которое сильно отличается от первоначального. Для наших целей будем использовать определение Т. Р. Грубера (T. R. Gruber), позже уточненное Р. Штудером (R. Studer): онтология – это явная и формальная спецификация концептуализации.

Онтология формально описывает некоторую предметную область. Обычно онтология состоит из конечного списка терминов и отношений между этими терминами. Термины обозначают важные понятия – *концепты* (*классы* объектов) рассматриваемой предметной области. Например, при описании университета в качестве концептов онтологии могут выступать сотрудники, студенты, курсы, лекционные аудитории, дисциплины.

К отношениям онтологии относят, как правило, иерархическое отношение классов. Иерархическое отношение «класс–подкласс» определяется следующим образом: класс C является подклассом класса C', если каждый объект класса C также является объектом класса C'. Например, все преподаватели являются сотрудниками. На рис. 1.3 изображена иерархия концептов для онтологии, описывающей деятельность университета.

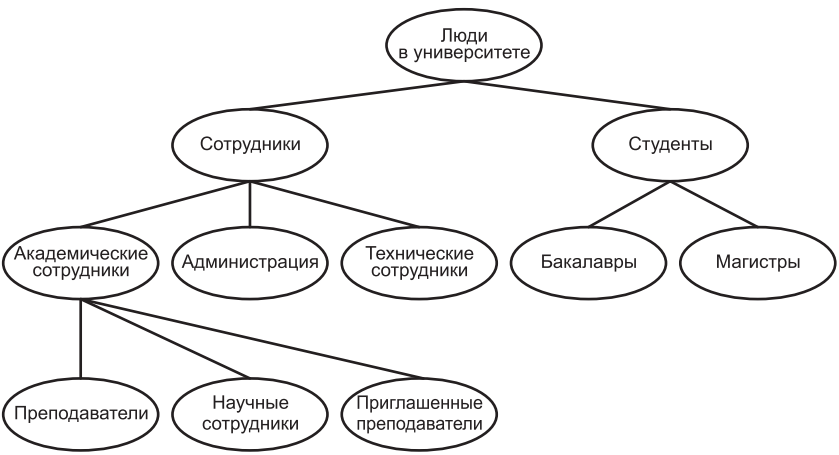

Рис. 1.3 ❖ Иерархия

Помимо отношений «класс–подкласс», онтология может включать в себя следующие типы информации:

- свойства[1] (например, свойство «преподает»: X преподает Y);
- ограничения на значения (например, только преподаватели могут преподавать курсы);
- взаимоисключающие утверждения (например, сотрудник университета не может быть одновременно и преподавателем, и техническим работником);
- характеристики логических связей между объектами (например, на каждой кафедре работает, по крайней мере, десять преподавателей).

В вебе онтологии используются для обеспечения *общего понимания предметной области*. Такое общее понимание необходимо для решения проблемы терминологического разнообразия. Например, в одном приложении для обозначения почтового индекса используется термин «Почтовый индекс», а в другом – просто «Индекс». Еще одна проблема заключается в том, что два приложения могут использовать один тот же термин, но с разными значениями. Например, в университете A под термином «курс» может пониматься программа подготовки бакалавров в целом (курс «Компьютерные науки»), а в университете B – конкретная дисциплина (курс «Программирование на языке C#»). Такие различия могут быть преодолены путем сопоставления терминологии конкретного приложения и терминологии, используемой в некоторой общей онтологии рассматриваемой предметной области, либо путем определения прямого отображения между конкретными онтологиями. В любом случае, онтологии обеспечивают семантическую совместимость.

Онтологии можно использовать для организации навигации по веб-сайтам. На многих современных веб-сайтах в левой части страниц представлены термины верхнего уровня иерархии понятий. Пользователь может «кликнуть» на любой их них и выбрать нужную подкатегорию.

Онтологии также применяются для повышения точности поиска в вебе. Поисковые системы, вместо того чтобы выдавать пользователю все страницы, в которых встречаются заданные (как правило, неоднозначные) ключевые слова, могут искать только те страницы, которые содержат точные концепты онтологии. Таким образом, могут быть преодолены различия в терминологии веб-страниц и терминологии запросов.

Кроме того, поисковые машины могут использовать обобщенную/уточненную информацию на основе онтологии. Если не были най-

[1] Свойство в онтологии – это отношение между классами. – *Прим. перев.*

дены документы, соответствующие заданному запросу, то поисковая система может предложить пользователю более общий запрос. Даже возможно, что поисковая система заблаговременно выполнит такой запрос, чтобы уменьшить время реакции в случае, если пользователь принимает это предложение. Если же, наоборот, в ответ на запрос будет возвращено очень много страниц, то можно предложить пользователю варианты уточняющего запроса.

В области искусственного интеллекта (ИИ) уже давно разрабатываются и используются языки онтологий. Эти разработки можно использовать в качестве основы при исследованиях в области сематического веба. В настоящее время к наиболее значимым языкам онтологий относят следующие языки:

- RDF Schema – язык представления онтологий, позволяющий описывать свойства и классы RDF-ресурсов, иерархии свойств и классов, а также область определения (домен) и область значения (диапазон) для свойств;
- OWL – более выразительный язык представления онтологий, который тоже позволяет описывать свойства и классы, а также отношения между классами (например, непересекаемость), ограничения кардинальности[1] (например, «точно один»), отношения эквивалентности для классов и свойств, различные типы свойств, характеристики свойств (например, симметричность) и классы, задаваемые перечислением.

1.2.3. Логика

Логика – это дисциплина, изучающая принципы рассуждения; она восходит своими корнями к Аристотелю. Логика прежде всего предлагает *формальные языки* для представления знаний. Еще логика предоставляет *широко распространенную формальную семантику*: в большинстве логик смысл высказывания определяется без необходимости дополнительных пояснений. Часто мы говорим о декларативных знаниях: описываем то, что уже известно, не заботясь о том, каким образом получены эти знания.

И в-третьих, автоматизированные машины логического вывода из заданных знаний могут получить (выводить) новые знания, тем самым делая неявные знания явными. Такие машины логического вывода широко изучаются в области ИИ. Ниже приведен пример ло-

---

[1] Ограничения кардинальности также называют ограничениями мощности. Например, лекции по дисциплине могут читаться только одним преподавателем, или многодетная мать имеет трех и более детей. – *Прим. перев.*

гического вывода. Предположим, известно, что все профессора являются преподавателями, все преподаватели являются сотрудниками и что Михаил – профессор. В логике предикатов информация выражается следующим образом:

профессор(*X*) → *преподаватель*(*X*),
преподаватель(*X*) → *сотрудник*(*X*),
профессор(*Михаил*).

Тогда можно вывести следующее:

преподаватель(*Михаил*),
сотрудник(*Михаил*),
профессор(*X*) → *сотрудник*(*X*).

Заметим, что знания из данного примера представляют собой типичные знания, которые хранятся в онтологиях. Таким образом, логика может быть использована для того, чтобы получать неявные знания на основе явных онтологических знаний. Благодаря возможности применять логический вывод в онтологии можно также выявлять в онтологических знаниях непредусмотренные отношения и противоречивые утверждения.

Но логика является более общей моделью, чем онтология. Она может быть использована интеллектуальными агентами для принятия решений и выбора вариантов действий. Например, агент интернет-магазина может принять решение о предоставлении скидки клиенту, основанное на правиле

постоянныйКлиент(*X*) → *скидка*(*X*, 5%),

где постоянность клиентов определяется на основе данных, хранящихся в корпоративной базе данных.

В общем случае существует компромисс между выразительной мощностью и вычислительной эффективностью логик. Более выразительная логика является менее эффективной, с точки зрения поддержки вывода. Иногда в выразительных логиках вывод вообще не может быть завершен. К счастью, большинство знаний для семантического веба можно представить в относительно простой форме. Например, наши предыдущие примеры описывают *правила* вида: «Если условия, то вывод», – где условия и вывод являются простыми высказываниями, и для того чтобы сделать вывод о возможности применения данных правил, необходимо рассмотреть только конечное число объектов. Такое подмножество логики, называемое логикой Хорна, является разрешимым и поддерживается эффективными инструментами логического вывода.