

УДК 004.82(111.1)

Онтология как посредник между пользователем и информационной системой

Шнак М.А., студент

*Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Информатика и системы управления»*

*Научный руководитель: Смирнова Е.В., к.т.н., доцент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана*

evsmirnova@bmstu.ru

Понятие «онтология» и принципы ее построения

Международный консорциум Интернет-программистов WorldWideWebConsortium (W3C) создал новый стандарт, который строит мост между миром знаний об организации самых разнообразных систем. Этот стандарт описывает принципы создания тезаурусов, классификаций, рубрикаций и онтологий[1]. Эта статья посвящена принципам автоматизированного создания онтологий. Рассмотрим вначале само понятие «онтология».

Понятие онтология сейчас активно применяется в информатике и искусственном интеллекте [2], хотя этот термин пришел из философии, где он обозначал часть метафизики, как учения о всем сущем, о его наиболее общих философских категориях, таких как бытие, субстанция, причина, действие, явление. При этом философская онтология, как наука, претендовала на полное объяснение причин всех явлений.

В инженерии знаний под онтологией понимается детальное описание некоторой проблемной области, используемое для формального и декларативного определения концептуализации этой области. Часто онтологией называют базу знаний специального вида [3], которую можно разделять, отчуждать и самостоятельно использовать в рамках рассматриваемой предметной области.

Можно сказать, что онтология — это точная спецификация некоторой области, которая включает в себя словарь терминов этой области и множество логических связей (типа «элемент-класс», «часть-целое»), которые описывают, как эти термины соотносятся между собой. Есть работы, в которых отмечается, что при таком подходе понятие

онтологии сильно пересекается с уже давно принятым в информатике и лингвистике понятием тезауруса[4]. Однако, отличие онтологии от тезауруса заключается в том, что онтология позволяет представить понятия в таком виде, чтобы быть пригодными для машинной обработки. Нередко онтологии используются в качестве посредника между пользователем и информационной системой, например, они позволяют формализовать договоренности о терминологии между членами сообщества, например, между пользователями некоторого корпоративного хранилища данных[4]. В центре большинства онтологий находятся классы, которые описывают понятия предметной области [5].

Итак, на формальном уровне онтология — это система, состоящая из наборов понятий и утверждений об этих понятиях, на основе которых можно строить классы, объекты, отношения, функции и теории. Практически все модели онтологий содержат определенные концепты (понятия, классы), свойства концептов (атрибуты, роли), отношения между концептами (зависимости, функции) и дополнительные ограничения, которые определяются аксиомами. Концептом может быть описание задачи, функции, действия, стратегии, процесса соображения и т. п. То есть, разработка онтологии предметной области включает в себя этап создания иерархии классов, свойств, отношений между классами, их характеристик [6].

Онтологические системы строятся на основе следующих четырех принципов:

- формализации, т. е. описания объективных элементов действительности, в единых строго определенных образцах (терминах, моделях и др.);
- использования ограниченного количества базовых терминов (сущностей), на основе которых конструируются все остальные понятия;
- внутренней полноты и
- логической непротиворечивости.

В отличие от обычного словаря для онтологической системы характерно внутреннее единство, логическая взаимосвязь и непротиворечивость используемых понятий.

Модели онтологий и языки их написания

Языки RDF и OWL сейчас позиционируются, как языки описания веб-страниц для обработки их поисковыми системами и прочими интеллектуальными программными агентами [7]. Язык OWL предназначен для того, чтобы выразить сложные концептуальные структуры, используемые для создания метаданных и для поддержки инструментов логического вывода.

Модель RDF (ResourceDescriptionFramework) – это разработанная консорциумом Всемирной паутины модель для представления данных, в особенности — метаданных[12]. Является аналогом семантической сети и включает также в себя XML-формат для хранения RDF-данных. Базовыми понятиями в модели RDF являются ресурсы, отношения, также называемые предикатами и свойствами, и утверждения, представляющие собой триаду "ресурс-субъект - отношение - ресурс-объект или литерал". Такой «плюс» RDF-модели, как расширяемость, дает возможность дальнейшей интеграции, с одной стороны, и\или расширения редактора словарей SKOS(SimpleKnowledgeOrganizationSystem) в более сложные словари, в том числе, вплоть до создания интернет-ориентированных онтологий, называемых OWL (WebOntologyLanguage) онтологиями.

Онтологии OWL — это последовательности аксиом и фактов, а также ссылок на другие онтологии. Они содержат компоненту для записи авторства и другой подробной информации, являются документами Web, на них можно ссылаться через URI[12].

Редакторы онтологий: плюсы и минусы

Рассмотрим среды, используемые в настоящее время для написания онтологий. Рассмотрим здесь три наиболее популярных средства создания и редактирования онтологий: редактор онтологий Стэнфордского университета "Protege" [6], который также позиционируется, как редактор баз знаний.

Плагин с открытым исходным кодом для новичков OWL онтологии - редактор SKOS (Simple Knowledge Organization System), он поддерживает авторские словари SKOS. Программный продукт SKOS имеет сопроводительный интерфейс написания приложений SKOS API, написанный на языке Java, который можно использовать для создания приложений на основе SKOS. Редактор онтологий SKOS предназначен для обеспечения наследования схем концепции, доступных для семантических веб-приложений, причем, более простых, чем онтологии на языке OntologyWebLanguage (OWL). Надо отметить, что тема разработки классификаций с помощью SKOS была разработана IBM [8].

Пример практического применения онтологии в университете

В качестве примера применения онтологий можно привести реализацию веб-представления Стэнфордского университета на основе стандартизированной онтологии, за основу которой можно взять систему Protégé [6]. Использование готовой онтологии позволяет разработчикам приступить к заполнению данных и построению шаблонов онтологий. В случае открытой доступа к RDF-данным возможна программная реализация

для поиска этих данных (к примеру, с помощью специальных запросов системы Google), собранных в едином хранилище и их предоставление пользователям (например, абитуриентом) в едином интерфейсе со специфическими функциями. Могут быть просто интегрированы данные факультетов и специальностей вуза, которые редактируются редактором онтологий на сервере основного веб-сайта вуза. В случае интеграции больших и часто меняющихся распределенных данных (например, для агрегации информации о конференциях в регионах с веб-узлов вузов и научных организаций), возможно использование RDF-хранилищ с открытыми интерфейсами для выборки только необходимых данных, например, как это сделано в [9]).

Заключение

Несмотря на то, что сегодня многие исследователи [11] считают, что построение «полезных» веб-онтологий является затратным с точки зрения опыта и финансов, что во многих случаях этот тип усилий (создание онтологий) может быть излишним, автор берет на себя смелость утверждать, что представляется перспективным и интересным рассмотрение возможностей использования OWL онтологий для представления предметных областей информационных веб-ресурсов - учебных, научных, познавательных, технических. В частности, OWL онтология была бы полезной для реализации системы контроля усвоения знаний студентами [13, 14]. При этом проще использовать уже существующие шаблоны, такие как составная часть семейства стандартизованных семантических сетей SKOS, назовем их «лучшим выбором».

Список литературы

1. Бархатов А.В. Разработка и интеграция веб-ресурсов с использованием онтологий. Режим доступа: http://www.sbras.ru/ws/show_abstract.dhtml?ru+115+8452 (дата обращения 14.05.2014)
2. Лапшин В. А. Онтологии в компьютерных системах. М.: Научный мир, 2010. 224 с.
3. Гладун А.Я., Рогушина Ю.В. Онтологии в корпоративных системах. Часть 1. Режим доступа: <http://www.management.com.ua/ims/ims115.html> (дата обращения 14.05.2014).
4. Добров Б.В., Иванов В.В., Лукашевич Н.В., Соловьев В.Д. Онтологии и тезаурусы: модели, инструменты, приложения. М.: Интернет-университет информационных технологий, 2009. 176 с.
5. Protégé, как редактор онтологий предметной области стэнфордского университета. Режим доступа: <http://protege.stanford.edu/> (дата обращения 14.05.2014).

6. Губанов Д.А., Макаренко А.В., Новиков Д.А. Методы анализа терминологической структуры предметной области (на примере методологии) // Управление большими системами. Вып.43, 2013. С. 5 – 33.
7. Классификация предметов использующих системы DITA и SKOS компании IBMdeveloperWorks. Режим доступа: <http://archive.is/MAyO> (дата обращения 14.05.2014).
8. Joseki RDF Server Режим доступа: http://www.cs.jyu.fi/ai/vagan/SW_Tools.ppt (дата обращения 14.05.2014).
9. Язык очередей для системы RDF. Andy Seaborne. RDQL - A Query Language for RDF. 2004. Режим доступа: <http://www.w3.org/Submission/RDQL/> (дата обращения: 14.05.2014)
10. Козлов И.А., Маслов И.А. Анализ и классификация нереляционных баз данных. Наука и образование. МГТУ им. Н.Э. Баумана. Электронный журнал. 2013. № 2.
Режим доступа: <http://sntbul.bmstu.ru/doc/552121.html> (дата обращения 14.05.2014)
11. Продукт компании Worldwidewebconcorium"W3 RDFPRIMER"
Режим доступа: <http://www.webcitation.org/686waxO2I> (дата обращения 14.05.2014)
12. Smirnova E.V. (2012) IBM, IMS and Mainframe: Education Quality System // Technical Symposium (Königstein, Germany 12-15 November 2012).
13. Козлов И.А., Смирнова Е.В., Шайхутдинов А.А., Шайхутдинов А.А. Программы оценивания качества образования // Труды Международных научно-методической конференции <Информатизация инженерного образования> ИНФОРИНО-2012 (Москва, 10-11 апреля 2012г.) - М.:МЭИ, 2012. С. 59-62