

Онтологический инжиниринг – это процесс проектирования и разработки онтологий. При отсутствии общепринятой методологии и технологии этот процесс не является тривиальной задачей. Он требует от разработчиков профессионального владения технологиями инженерий знаний – от методов извлечения знаний до структурирования и формализации.

Такую работу для интеллектуальных систем обычно выполняют инженеры по знаниям (knowledge engineers). Сегодня фактически ни один российский вуз не готовит аналитиков или инженеров по знаниям, владеющих методами инженерии знаний. Наиболее близкими являются специальности инженера-системотехника и специалиста по информационным технологиям (ИТ). На наш взгляд целесообразно включать в учебный процесс примеры построения онтологий в заданных предметных областях. В данной статье представлено описание создания онтологии студентами в рамках курсовых работ по специальным дисциплинам. Данная онтология принадлежит к сфере управления механизацией и обеспечением производства на примере одного из сельскохозяйственных предприятий Белгородской области.

Онтологии призваны создать общее представление об объекте исследования через разработку концепции его модели на категориальном уровне. Преимущества информационных технологий онтологий достигаются за счет преобразования данных на языке категорий в ценные, смысловые руководства к действию.

В данном аспекте, онтология является попыткой всеобъемлющей и детальной формализации некоторой области знаний с помощью концептуальной схемы. Обычно такая схема состоит из иерархической структуры данных, содержащей все релевантные классы объектов, их связи и правила (теоремы, ограничения), принятые в этой области.

Принято считать что, не существует единственного «правильного» способа или методологии разработки онтологий. Обычно используется итеративный подход к разработке онтологий.

Существуют традиционные языки спецификации онтологий (Ontolingua, CycL, языки, основанные на дескриптивных логиках, такие как LOOM, и языки, основанные на фреймах - OKBC, OCML, Flogic). Более поздние языки, основанные на Web-стандартах, такие как XOL, SHOE, RDF(S), DAML+OIL, OWL созданы специально для обмена онтологиями через Web.

На наш взгляд можно считать целесообразным использование системы Protégé_3.2.1, так как она является наиболее доступной, бесплатной, имеет интуитивно понятный интерфейс, встроенные примеры готовых онтологий, а также поддерживает экспорт онтологий в самые популярные форматы их представления (HTML, XML, RDF, OWL). Данные форматы позволяют использовать онтологию для категоризации web-страниц и в интеллектуальных программных агентах.

В данном случае областью проектируемой онтологии являлась классификация сельскохозяйственной техники и оборудования [1, с. 206]. В онтологию следует включить понятия, описывающие различные типы сельскохозяйственной техники, основные виды оборудования и понятие их агрегируемости.

Процесс проектирования онтологий можно разделить на несколько этапов. На начальном этапе проектирования онтологии целесообразно составить полный список всех терминов, которые будут в ней использованы, не беспокоясь о пересечении понятий, которые они представляют, об отношениях между терминами, о возможных свойствах понятий или о том, чем являются понятия – классами или слотами.

Следующие два этапа построения онтологии – разработка иерархии классов и определение свойств понятий (слотов) – тесно переплетены. Обычно в иерархии дается несколько формулировок понятий, и затем описываем свойства этих понятий и т.д. Эти два шага – самые важные шаги в процессе проектирования онтологий.

Существуют следующие подходы для разработки иерархии классов:

1. Процесс нисходящей разработки начинается с определения самых общих понятий предметной области с последующей конкретизацией понятий. 2. Процесс восходящей разработки начинается с определения самых конкретных классов, листьев иерархии, с последующей группировкой этих классов в более общие понятия. 3. Процесс комбинированной разработки – это сочетание нисходящего и восходящего подходов: сначала определяются более заметные понятия, а затем они соответствующим образом обобщаются и ограничиваются.

На рисунке 1 показано деление на различные уровни обобщения.

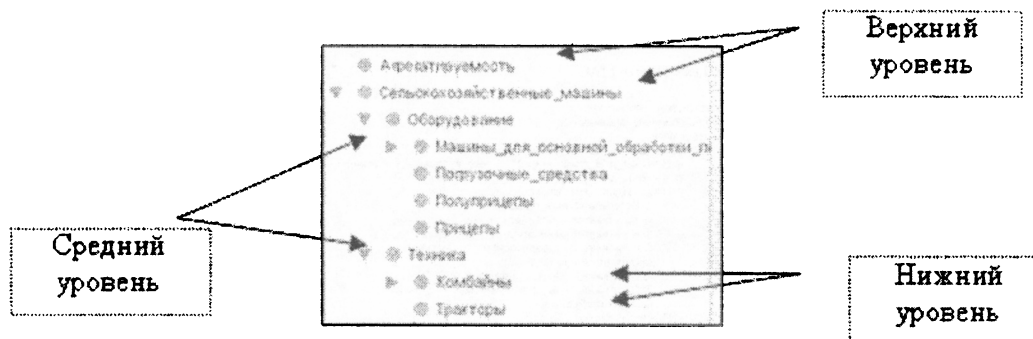


Рис. 1. Уровни таксономии Сельскохозяйственные машины

Если класс А – надкласс класса В, то каждый экземпляр В также является экземпляром А. Другими словами, класс В представляет собой понятие, которое является «разновидностью» А.

Например, каждая сеялка – обязательно относится к оборудованию. Поэтому класс «Сеялки» – подкласс класса «Оборудование». Для каждого свойства из списка нужно определить, какой класс оно описывает. Эти свойства станут слотами, привязанными к классам.

Все подклассы класса наследуют слот этого класса. Слот должен быть привязан к самому общему классу, у которого есть данное свойство.

Слоты могут иметь различные facets, которые описывают тип значения, разрешенные значения, число значений (мощность) и другие свойства значений, которые может принимать слот.

При определении слотов можно следовать следующим правилам:

- если список классов, определяющих диапазон значений слота или домен, включает класс и его подкласс, то подкласс следует удалить;
- если список классов, определяющих диапазон значений слота или домен, включает все подклассы класса А, но не включает сам класс А, то в диапазон значений входит только класс А [2, с. 20].

Завершающим этапом является создание отдельных экземпляров классов в иерархии. Для этого требуется: 1) выбрать класс; 2) создать отдельный экземпляр этого класса; 3) ввести значения слотов.

Таким образом, можно выделить некоторые фундаментальные правила разработки онтологии: 1) не существует единственного правильного способа моделирования предметной области – всегда существуют жизнеспособные альтернативы, поэтому лучшее решение почти всегда зависит от предполагаемого приложения и ожидаемых расширений; 2) разработка онтологии – это обязательно итеративный процесс; 3) понятия в онтологии должны быть близки к объектам (физическим или логическим) и отношениям в интересующей предметной области – это существительные (объекты) или глаголы (отношения), которые описывают рассматриваемую предметную область [3, с. 23].

В нашем случае система Protégé также позволяет создавать слоты, которые могут быть использованы для описания отношений между классами, которые не определены в иерархии классов. Для этого существуют facets типа Instance (экземпляр) или Class (класс). Закладка запросов (Queries) позволяет получать сведения из проекта по всем экземплярам классов, которые удовлетворяют интересующим критериям.

Наиболее распространенными и востребованными являются онтологии в форматах RDF, OWL и HTML. Готовую онтологию можно экспортировать в перечисленные форматы.

Иерархия классов и описание экземпляра класса в формате HTML представлена на рисунке 2.



Дата: 04:21:2008, 14:54:25 PM, Moscow Daylight Time

Система Protégé разработана на кафедре Информатики, периодизация © 1994-2008 Кирилл Чибриков

Рис. 2. Иерархия классов онтологии в формате HTML

Делая вывод, можно заметить, что применение онтологий позволяет повысить эффективность освоения студентами новой дисциплины за счет четкого структурирования системы понятийных знаний предметной области. А освоение и самостоятельное построение онтологий студентами позволяет формировать системное понятийное мышление.

Литература

1. Козлов В.В. Справочник инженера-механика агропромышленного производства. — М.: ФГНУ «Росинформагротех», 2000.- 384 с.
2. Муромцев. Д.И. Онтологический инжиниринг знаний в системе Protégé. -- СПб: СПб ГУ ИТМО, 2007. – 62 с.
3. Пивоварова Л.М., Рубашкин В.Ш. Компоненты онтологических систем и их реализация в современных проектах X Всероссийская объединенная конференция "Интернет и современное общество" (IMS-2007). – СПб: СПб ГУ. – 25с.

Коднер Г.Я.

ИСПОЛЬЗОВАНИЕ ОНТОЛОГИИ ДЛЯ СОЗДАНИЯ ЭЛЕКТРОННЫХ БИБЛИОТЕК

НОУ Гуманитарный университет

Онтология – раздел философии, учение о бытии, в котором исследуются всеобщие основы, принципы бытия, его структура и закономерности; термин введен немецким философом Р. Гоклениусом.

Онтология (в информатике) — это попытка всеобъемлющей и детальной формализации некоторой области знаний с помощью концептуальной схемы. Обычно такая схема состоит из иерархической структуры данных, содержащей все релевантные классы объектов, их связи и правила (теоремы, ограничения), принятые в этой области. Этот термин в информатике является производным от древнего философского понятия «онтология».

Одно из основных направлений информатизации образования является создание электронных библиотек.

Основной задачей, возникающей при работе с полнотекстовыми базами данных, является задача поиска документов по их содержанию. Однако, ставшие традиционными средства контекстного поиска по вхождению слов в документ, представленные, в частности, поисковыми машинами в Интернет, зачастую не обеспечивают адекватного выбора информации по запросу пользователя.

Основная проблема заключается в сложности точной формулировки запроса – подбора ключевых слов, которые предстоит искать в телах документов. Это может быть связано с рядом причин: недостаточным знанием пользователем терминологии предметной области, наличием в языке многозначных и синонимичных слов, и даже орфографическими ошибками в написании искомых слов, которые могут встречаться как в текстах, так и в самом запросе.

Другая фундаментальная причина заключается в том, что иногда пользователь не знает точно, какую именно информацию ему хотелось бы получить, имея лишь общее представление о границах своих интересов. Так, например, пытаясь расширить свои познания в области компьютерной лингвистики, на поисковом сервере Google вы просто получите список из сотен тысяч документов, содержащих слова “computer” и “linguistic”. А