

МОДЕЛИРОВАНИЕ БАЗ ЗНАНИЙ В РЕДАКТОРЕ ОНТОЛОГИЙ PROTÉGÉ

В.С. Оскерко, З.В. Пунчик

*УО «Белорусский государственный экономический университет»,
г. Минск, Республика Беларусь*

Важнейшей ценностью в современном информационном обществе являются знания. Для их получения необходимы новые эффективные технологии обработки информации и преобразования ее в нужные человеку знания.

Во многих случаях задачи, решаемые на практике в различных предметных областях, могут быть классифицированы как задачи нечеткой логики ввиду отсутствия полной и достоверной информации для принятия эффективного решения. Они характеризуются неполнотой, неоднозначностью, неопределенностью и зачастую противоречивостью исходной информации. В таких случаях требуется осуществить формализованное описание предметной области и сформировать ее базу знаний для последующего использования этой базы при решении возникающих практических задач.

Существуют различные программные средства для создания баз знаний. В настоящее время популярным является редактор онтологий Protégé, который имеет богатые функциональные возможности и свободно распространяется, что делает его привлекательным для учебного процесса в дисциплинах, изучающих технологии баз знаний.

Онтология – это формальное явное описание понятий в рассматриваемой предметной области и отношений между ними. Онтология представляет предметную область в реальном мире и понятия в ней также должны отражать эту реальность. Ее разработка дает возможность:

- *общего понимания специалистами структуры информации предметной области.* К примеру, пусть несколько различных веб-сайтов содержат информацию по экономике. Если эти веб-сайты совместно используют и публикуют одну и ту же базовую онтологию терминов, которыми они все пользуются, то программные агенты могут извлекать информацию из этих различных сайтов, накапливать ее и использовать накопленную информацию для ответов на запросы пользователей или как входные данные для других приложений;

- *повторного использования знаний предметной области.* Например, для IT-сферы необходимо сформулировать понятие «персональный компьютер». Если одна группа ученых детально разработает такую онтологию, то другие могут просто повторно использовать ее;

- *явного определения допущений в предметной области,* что дает возможность легко изменить эти допущения при изменении наших знаний о предметной области. Кроме того, явные спецификации знаний в предметной области полезны для новых пользователей, которые должны узнать значения терминов предметной области;

- *анализа и использования знаний.* Формальный анализ терминов чрезвычайно ценен как при попытке повторно использовать существующих онтологий, так и при их расширении [1, с. 6–7].

Онтология определяет общий словарь для людей, которым нужно совместно использовать информацию в предметной области. Единого универсального подхода к созданию онтологий, который бы однозначно привел к успешному результату, не существует. Для любой предметной области может существовать бесчисленное количество онтологий. Каждая новая онтология – это всего лишь один из способов описания модели предметной области. Выбор лучшей онтологии осуществляется исходя из того, которая из них будет эффективнее всего работать для намеченной задачи.

Процесс разработки онтологий является итеративным. При создании онтологии нужно стремиться к максимальному близкому отображению объектов и отношений между ними в избранной предметной области. При правильном моделировании онтология может быть представлена предложениями, где вероятней всего в качестве существительных выступают объекты, а отношений – глаголы. Для работы с онтологиями существуют программные продукты, называемые редакторами онтологий. Популярным среди них является редактор онтологий PROTÉGÉ, разработанный в Стэнфордском университете США. Его платформу-независимая расширяемая среда и удобный графический интерфейс дают пользователю возможность быстро приступить к созданию своих онтологий. Имеющиеся в нем развитые и простые в управлении средства позволяют визуализировать представления модели предметной области.

Разработка онтологии включает:

- выделение предметной области онтологии;
- определение объектов предметной области и их связей;
- организацию классов (объектов) в некоторую иерархию: базовый класс → подкласс;
- формирование фреймов для описания классов и подклассов через определение слотов (свойств классов);
- заполнение слотов значениями (создание экземпляров классов);
- графическое представление онтологии.

Редактор онтологий Protégé позволяет разрабатывать онтологии как семантические сети и на их основе создавать базы знаний по фреймовой модели, а также формировать различные пользовательские запросы к этим базам с целью удовлетворения своих информационных потребностей.

В хорошо спроектированной онтологии все классы, имеющие одно и то же более общее понятие, должны находиться на одном уровне иерархии.

Онтология вместе с множеством экземпляров классов составляет базу знаний, к которой можно «посылать» различные запросы.

Владея работой с базой знаний, специалисты любых предметных областей могут решать разнообразные практические задачи с нечеткой логикой. Освоение технологии моделирования базы знаний на основе онтологии упрощается при использовании учебного примера онтологии предметной области, доступного для понимания и позволяющего в максимальной степени проиллюстрировать функциональные возможности избранного редактора онтологий.

Авторами для разработки учебного примера онтологии выбрана предметная область «Университет», включающая представление о структурных подразделениях университета и их характеристике: руководителе, номере телефона, месторасположении, времени работы, количестве работников, а также о руководстве ими ректором. Ее упрощенная структура приведена на рисунке 1.

На базе данного примера разработаны на алгоритмическом уровне методические рекомендации:

- по созданию и сохранению проекта «Университет»;
- созданию абстрактных (не имеющих экземпляров) и конкретных классов, а также их иерархии наследования: добавлению новых классов, удалению существующих, перемещению классов с одного уровня на другой, переименованию классов. Для проекта «Университет» созданная в среде Protégé иерархия классов выглядит так, как представлено на рисунке 2;

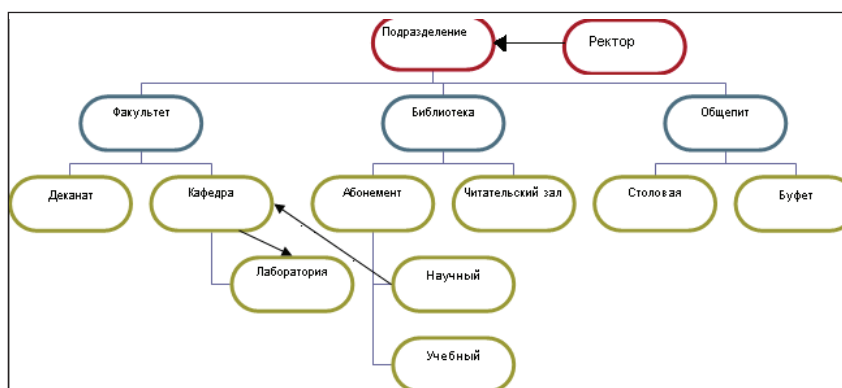


Рисунок 1 – Структура предметной области «Университет»

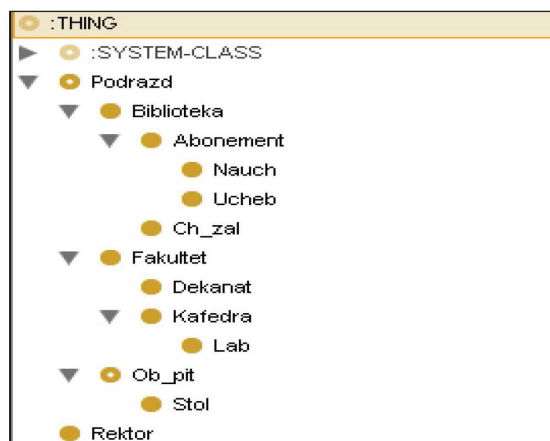


Рисунок 2 – Иерархия классов проекта «Университет» в среде Protégé

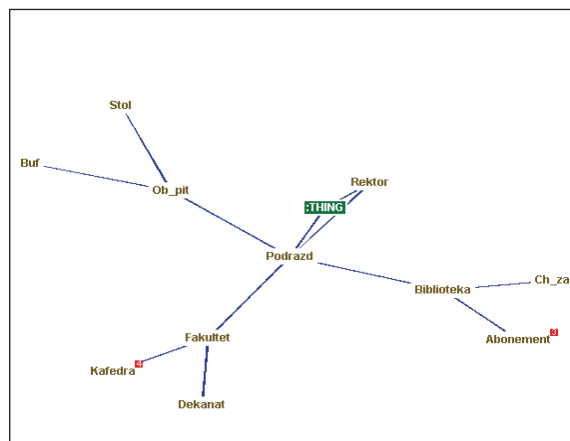


Рисунок 3 – Графическое представление онтологии проекта «Университет»

- созданию слотов с различными типами значений (String, Integer, Symbol и др.) и фасетами – ограничениями на значения, например, минимальное значение, максимальное значение, значение по умолчанию;
- связыванию слотов с классами;
- созданию связи определенного типа между классами посредством специальных слотов;
- созданию экземпляров классов и использованию для этого стандартных и специальных форм, в которых пользователь может изменить расположение и размер их элементов, а также скрыть некоторые из них;
- созданию связи между классами на уровне экземпляров;
- графическому представлению онтологии на разных уровнях иерархии классов (в том числе с изображением связей между классами, экземпляров классов), ее редактированию и оформлению с помощью TGvizTab – одного из многих инструментальных средств, подключаемых к Protégé. Пример графического представления онтологии проекта «Университет» для третьего уровня приведен на рисунке 3;
- созданию запросов, извлекающих из базы знаний различную информацию о предметной области «Университет», в которых в условиях отбора используются различные операции сравнения, шаблоны, логические операции, в том числе вложенных запросов;
- сохранению и загрузке сохраненных запросов из библиотеки запросов;
- запуску запросов.

Апробация разработанных методических рекомендаций осуществлена в Белорусском государственном экономическом университете при изучении темы «Базы знаний и модели представления знаний» раздела «Технологии баз данных и знаний» дисциплины «Компьютерные информационные технологии» на лабораторных занятиях. Для закрепления практических навыков по моделированию базы знаний на основе онтологии в среде Protégé, полученных на лабораторных занятиях, к методическим рекомендациям добавлен комплекс заданий для самостоятельной работы.

Апробация показала, что, освоив моделирование баз знаний с помощью онтологии, будущие специалисты приобретают навыки самостоятельной разработки баз знаний с моделью представления знаний в виде семантической сети в различных предметных областях. На основе созданных онтологий они смогут формализовывать накопленные специалистами знания, многократно их использовать, передавать другим, решать разнообразные практические задачи с использованием разработанных баз знаний. Владение технологией моделирования баз знаний значительно расширяет информационную и профессиональную компетентность подготавливаемых специалистов.

ЛИТЕРАТУРА

1. Муромцев, Д.И. Онтологический инжиниринг знаний в системе Protégé / Д.И. Муромцев. – СПб.: СПб ГУ ИТМО, 2007. – 63 с.