Лабораторная работа № 7

Построение онтологии в системе Protégé

**Цель**: Получение опыта в анализе предметной области для построения онтологии на основе фреймовой модели в системе Protégé.

**Тема:**Получить представление о кодировании знаний в виде онтологий в стандарте RDF в редакторе онтологий Protégé. 

**Ход работы**:

* Скачать редактор [ProtégéFrames](http://protege.stanford.edu/) (3.4.x)
* Прочитать *Муромцев Д.И.* [Онтологический инжиниринг знаний в системе Protégé](http://window.edu.ru/window_catalog/redir?id=54429&file=itmo240.pdf): Методическое пособие. — СПб: СПбГУ ИТМО, 2007.
* Разобрать там пример онтологии редакции газеты.
* Закодировать самому подобную онтологию по заданию.

**Критерии оценки Онтологии:**

* **Ясность**: онтология должна быть ясной и легко передавать подразумеваемый смысл. Она должна быть объективной;
* **Последовательность**: в ней должны содержаться утверждения, которые не противоречат друг другу, иерархии понятий, связывающим их отношениям, экземплярам.
* **Возможность расширения**: наличие возможности введения новых элементов без пересмотра остальных элементов;
* **Минимальная степень специализации онтологии**: нежелательность полного подчинения онтологии конкретной задаче, что может осложнить ее последующее использование в других задачах.

Онтологии

В значение слова "онтология" мы включаем каталог терминов (понятий), используемых в предметной области; правила говорящие о том, как эти понятия могут быть скомбинированы и размещены в этом домене, чтобы отразить адекватное описание предметной области; и «санкционированные выводы», которые могут быть сделаны, когда такие утверждения использованы в домене.

В каждом домене, наблюдается ситуация такая, что люди выделяют в этой предметной области: (концептуальные или физические) объекты, зависимости, и местоположения. Через различные языковые механизмы, мы присоединяем определенные описатели (такие как, названия, имена существительные и т.д.) к этой ситуации. В контексте «онтологии», мы используем термин «отношение», чтобы ссылаться к определенному описателю, который, в свою очередь, ассоциирован с зависимостью в реальном мире. Мы используем термин "понятие", чтобы обратиться к определенному описателю, который относится к объекту или вещи в реальном мире. В онтологии, мы пробуем каталогизировать описатели (подобно словарю данных) и создать модель домена, описанного этими описателями.

Таким образом, при формировании онтологии необходимо сделать три вещи. (1) каталогизировать понятия, (2) фиксировать ограничения, которые управляют тем, как эти понятия могут использоваться, чтобы делать описательные утверждения относительно домена, (3) и затем сформировать модель, которая, когда обеспечено определенное описательное утверждение, может генерировать «соответствующие» дополнительные описательные утверждения.

Любой домен с определенной предметной областью имеет свою собственную терминологию, отличительный словарь, который используется, чтобы говорить относительно характерных объектов и процессов, которые образуют домен. Библиотека, например, вовлекает в ее собственный словарь, понятия имеющие отношение к книгам, ссылкам, библиографиям, журналам, и т.д. Точно так же производство полупроводника имеет собственный язык - чипы, подложки, проекты, и так далее. Характер данного домена, таким образом отражается в языке, используемом для общения в этой предметной области.

**Разработка онтологий**

Онтологии необходимо разрабатывать затем, что бы:

* Разделять общее понимание о структуре информации между людьми и между программными агентами.
* Получить возможность повторного использования знания и той или иной предметной области. Это оберегает от изобретения велосипедов, а так же вводит стандарты в данной области, обеспечивающие совместимость.
* Чтобы явнозаписать, что мы знаем о какой-то области
* Чтобы отделитьзнания об области от процедур. Дает повторное использование знаний и повторное использование процедур.

Разработка онтологий находится на более высоком уровне абстракции, чем ООП проектирование.

|  |  |
| --- | --- |
| **Разработка онтологий** | **ООП проектирование** |
| *Отражает строение мира* | *Отражает построение кода и данных* |
| *фокусируется на структуре концептов* | *фокусируется на поведении (методы) и передаче сообщений между объектами* |
| *Физическое представление не важно* | *описывает физическое представление данных (int, char, etc.)* |

Процесс разработки онтологий

Разработка онтологий – это творческий итеративный процесс; тем не менее, каждая онтология отражает структуру реального мира и может быть объективно оценена как хорошо или плохо отражающая всю полноту знаний.

1. Не существует единственного правильного способа моделирования предметной области – всегда существуют жизнеспособные альтернативы. Лучшее решение почти всегда зависит от предполагаемого приложения и ожидаемых расширений. Так же, та онтология лучше – которая короче.
2. Разработка онтологии – это обязательно итеративный процесс. Под итеративным процессом понимается неоднократный проход по онтологии с целью ее уточнения, то есть на начальном этапе строится черновой вариант. Затем мы проверяем и уточняем составленную онтологию, добавляя детали, возможно, частично или даже полностью пересматривая начальную онтологию.
3. Элементы онтологии должны быть близки к объектам (физическим или логическим) и отношениям в интересующей вас предметной области. Скорее всего, они соответствуют существительным (объекты) или глаголам (отношения) в предложениях, которые описывают вашу предметную область.

Построение онтологий может быть в общих чертах описано при помощи следующего итеративного процесса, в котором каждые из нескольких этапов могут выполняться параллельно и по несколько раз.

**Шаг 1. Определение области и масштаба онтологии**

* Какую область будет охватывать онтология?
* Для чего мы собираемся использовать онтологию?
* На какие типы вопросов должна давать ответы информация в онтологии?
* Кто будет использовать и поддерживать онтологию?
* Онтология не должна содержать ВСЕ возможные знания о домене
  + не требуется уточнять или обобщать более, чем необходимо
  + не требуется включать все возможные свойства классов

**Шаг 2. Рассмотрение вариантов повторного использования существующих онтологий**

* Библиотекионтологий
  + DAML ontology library (www.daml.org/ontologies)
  + Protégé ontology library (protege.stanford.edu/plugins.html)
* Онтологииверхнегоуровня
  + IEEE Standard Upper Ontology (suo.ieee.org)
  + Cyc (www.cyc.com)
* Общие онтологии
  + DMOZ (www.dmoz.org)
  + WordNet (www.cogsci.princeton.edu/ wn/)
* Специальные онтологии
  + UMLS SemanticNet
  + GO (Gene Ontology) (www.geneontology.org)
* Библиотека онтологий Ontolingua (<http://www.ksl.stanford.edu/software/ontolingua/>)
* Общедоступныекоммерческиеонтологии (например, UNSPSC (www.unspsc.org), RosettaNet (www.rosettanet.org), DMOZ ([www.dmoz.org](http://www.dmoz.org/)))

**Шаг 3. Перечисление важных терминов в онтологии**

* Какие термины являются ключевыми
* Каковы свойства этих понятий
* Что мы хотим сказать (выразить) этими термами?
* неформально сгруппировать
* перефразироватьи уточнить термины, чтобы получить неформальные определения концептов
* нарисовать неформальную диаграмму

**Шаг 4. Определение классов и иерархии классов**

* Нисходящая разработка
  1. Начиная с общих понятий
* Восходящая разработка
  1. Движение от конкретных классов
* Комбинированная разработка
* Организация классов в таксономию (установка отношений наследования)
  1. Класс может иметь множество суперклассов (множественное наследование)
  2. Все родственные понятия в иерархии классов должны быть на одном уровне иерархии
  3. Имена классов должны быть или все в единственном числе или все в множественном числе (Animalisnotakind-ofAnimals).
* Документация классов

**Шаг 5. Определение свойств классов – слотов**

* Подкласс наследует все слоты от всех суперклассов.

**Шаг 6. Определение фацетов слотов (ограничений на слоты)**

* Тип значений
  + Строка
  + Число
  + Булево значение
  + Нумерованные слоты (один из множества)
  + Слоты-экземпляры – ссылка на другой класс
* Количество значений (кардинальность)
* Свойства значений
  + Минимум, максимум
  + Значение по умолчанию

**Шаг 7. Создание экземпляров**

1. выбрать класс,
2. создать отдельный экземпляр этого класса
3. ввести значения слотов.

Средства разработки онтологий

* AltovaSemanticWorks
  + основана на представлении графов дескриптивных логик
* Ontolingua
  + веб-ориентированная
  + нацелена на совместную распределённую работу
* **Protégé**
  + две версии — основанная на графах дескриптивных логик и на фреймовом представлении

Задания

**Задание 1.** Онтология животных (и растений), упоминающихся в детских книжках, с целью создания предметного указателя. Должна включать:

* где живут
* что едят
* (хищное, травоядное, всеядное)
* опасны ли они
* большие ли они
* сказочные или нет
* немного анатомии(количество ног, крыльев, пальцев, и т.п.)

**Задание 2.** Онтология по компьютерным вирусам.

**Задание 3.** Онтология "Астрономия" – Разновидности звезд, Планет, Дополнительные объекты: Астероид, Комета, Макрообъекты: Туманность, Созвездие.

**Задание 4.** Онтология «Деканат».

**Задание 5.** Онтология «Персональный компьютер».

**Задание 6.** Онтология «Туристические места Беларуси».

**Задание 7.** Онтология «Семейные отношения».

**Задание 8.** Онтология «Пиццы». Какие бывают пиццы, из чего состоят и т.д.

**Задание 9.**Онтология "Наземный транспорт", Легковой автомобиль, Грузовик, Автобус, Мотоцикл, Велосипед.

**Задание 10.**Онтология "Электронная аппаратура",Телевизор, DVD-плейер, Компьютер, радиоприёмник, сотовый телефон, цифровой фотоаппарат.

**Задание 11.**Онтология "Кухонные приборы" - Микроволновая печь, Кухонный комбайн, Кофеварка, Посудомоечная машина, Миксер, Пароварка.

**Задание 12.**Онтология "Одежда", объекты – Юбка, брюки, платье, Костюм, Свитер, Пальто, Куртка.

**Задание 13.**Онтология "Обувь" – Туфли, Босоножки, Сапоги, Шлёпанцы, Ботинки.

**Задание 14.**Онтология "Олимпийские виды спорта".

**Задание 15.**Онтология "Инструменты" – Рубанок, Стамеска, Зубило, Пила, Отвёртка, Плоскогубцы, Кусачки, Струбцина.

**Задание 16.**Онтология " Воздушный транспорт " – Самолёт, Вертолёт, Планер, Ракета, Аэростат, Парашют.