Л. М. Пивоварова, В. Ш. Рубашкин

Компоненты онтологических систем и их реализация в современных проектах

Санкт-Петербургский государственный университет
Институт филологических исследований
Лаборатория информационных систем

Онтология – спецификация концептуализации. (Т. Грубер)

Данная работа является обзором основных точек зрения на представление знаний и управление знаниями в онтологических системах.

Существующие обзоры:

1. Овдей О. М., Проскудина Г. Ю. Обзор инструментов инженерии онтологий // Труды Шестой Всероссийской научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", 2004
2. Gomez-Perez A., Fernando-Lopez M., Corcho O. Ontological Engineering with examples from the areas of Knowledge management, e-Commerce and the Semantic Web – 2nd. Ed. – London, Springer-Verlag – 2004
3. Mizoguchi R. Ontology Engineering Environments // Handbook on Ontologies; ред. Staab S., Studer R. – Berlin, Springer-Verlag - 2003
4. L. Stojanovic, B. Motik Ontology Evolution within Ontology Editors // Proceedings of EON2002 Evaluation of Ontology-based Tools, 2002

Мы считаем наиболее важными следующие аспекты:

- Базовый язык представления знаний и аксиоматика онтологии.
- Наполнение онтологии знаниями и интеграция накопленных ранее знаний.
- Ограниченный логический вывод.
- Связь онтологии с естественным языком.
- Инструментарий для работы с онтологиями (онторедактор).

Существующие проекты

Protege
SUMO
KAON2
CYC
DOLCE
Hozo
AKT
Omega
WebODE
WebONTO
DIP

Protégé

- Онторедактор и инструмент для построения баз знаний
- Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine
- Protégé Frames / Protégé OWL

Язык представления: OWL

Соответственно, могут быть реализованы основные выразительные возможности OWL:

- *различение индивидов, классов и свойств*
- *иерархия классов и свойств*
- *ограничения на свойства*
- *области значений и области применения*
- *перечислимые классы*
- *взаимоисключающие классы*

Функции ввода и редактирования реализованы, по возможности, в виде отдельных кнопок и выпадающих списков.

pizza.owl Protégé 3.2.1 (file:\C:\Program%20Files\Protege_3.2.1\examples\pizza\pizza.owl.pprj, OWL / RDF Files)

File Edit Project OWL Code Tools Window Help

protégé

Metadata (pizza.owl) | OWLClasses | Properties | Individuals | Forms

SUBCLASS EXPLORER

For Project: pizza.owl

Asserted Hierarchy

- owl:Thing
  - DomainConcept
    - Country
    - IceCream
    - Pizza
      - CheeseyPizza
      - InterestingPizza
      - MeatyPizza
      - NamedPizza
      - NonVegetarianPizza
      - RealItalianPizza
      - SpicyPizza
      - SpicyPizzaEquivalent
      - VegetarianPizza
      - VegetarianPizzaEquivalent1
      - VegetarianPizzaEquivalent2
    - PizzaBase
    - PizzaTopping
  - ValuePartition
    - Spiciness

CLASS EDITOR

For Class: VegetarianPizza (instance of owl:Class) Inferred View

Annotations

Property	Value	Lang
rdfs:comment	Any pizza that does not have fish topping and does not have meat topping is a VegetarianPizza. Members of this class do not need to have any toppings at all.	en
rdfs:label	PizzaVegetariana	pt

Properties and Restrictions

- hasBase (single PizzaBase) (someValuesFrom PizzaBase)
  - PizzaBase [from Pizza]
- hasTopping (multiple PizzaTopping)

Defining Classes

- Pizza
- not (hasTopping some MeatTopping)
- not (hasTopping some FishTopping)

Disjoints

- NonVegetarianPizza

Logic View | Properties View

- позволяет быстро и относительно просто сконструировать небольшую предметную онтологию
- имеет расширяемую архитектуру, которая позволяет легко встраивать его в прикладные программы (за счет Java-плагинов)

- не имеет никакой встроенной онтологии и рассчитана на создание прикладных онтологий с нуля
- не предусмотрена связь с естественным языком

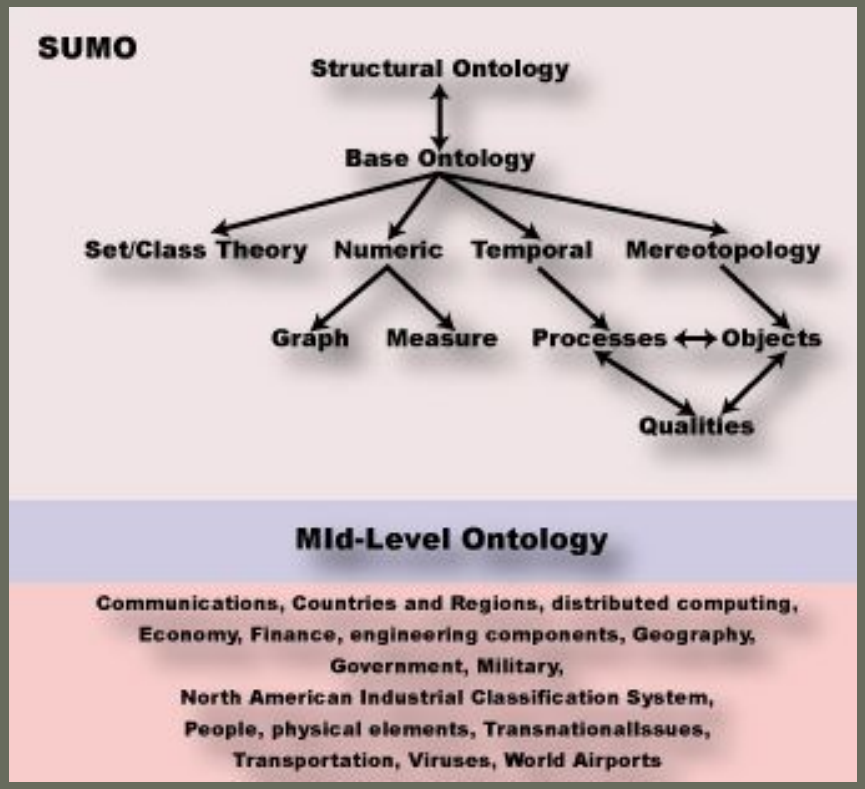
SUMO

- Suggested Upper Merged Ontology, проект IEEE

***Цель**: интеграция существующих онтологий в единую, всеобъемлющую, способную к расширению структуру, которая имела бы статус универсального стандарта и могла бы использоваться в различных прикладных и исследовательских проектах*

- SUMO – онтология верхнего уровня: ~1000 понятий, ~5000 аксиом
- MILO – онтология промежуточного уровня
- Отраслевые онтологии
- Весь корпус онтологий:

  ~20000 понятий, ~60000 аксиом



- Язык представления: SUO-KIF; для SUMO выполнен перевод на OWL
- Разработана полная система соответствий понятий SUMO с лексическими единицами WordNET (для английского языка)
- Онторедактор и машина логического вывода отсутствуют

KAON2

- Инфраструктура для управления онтологиями, ориентированная на SemanticWeb и бизнес-приложения
- University of Karlsruhe, University of Manchester

- Обладает машиной логического вывода и ответов на запросы
- Позволяет управлять онтологиями
- Имеет программный интерфейс для взаимодействия с другими приложениями
- Может использоваться в качестве онтологического сервера
- Обладает механизмом извлечения знаний из реляционных баз данных

- Онтология рассматривается с точки зрения структуры, а не содержания
- Машина логического вывода наилучшим образом работает с онтологиями с компактной T-частью и обширной A-частью
- Позволяет импортировать онтологии на языках OWL-DL, SWRL и F-Logic;

  для задач логического вывода используется собственный внутренний язык, основанный на клаузах Хорна

CYC

- Многоцелевая база знаний и машина логического вывода, направленная на семантизацию программного обеспечения за счет базовых интуитивных знаний (здравого смысла)
- Cycorp
- Коммерческая система, ориентированная на использование в бизнес-процессах

- Язык представления знаний: CycL - гибридный язык, сочетающий в себе свойства фреймов и логики предикатов
- Основное содержание онтологии – это предложения, сконструированные из предикатов
- Многоуровневое представление онтологии в виде иерархии микротеорий
- Не предъявляется требование непротиворечивости на глобальном уровне

328 000 концептов

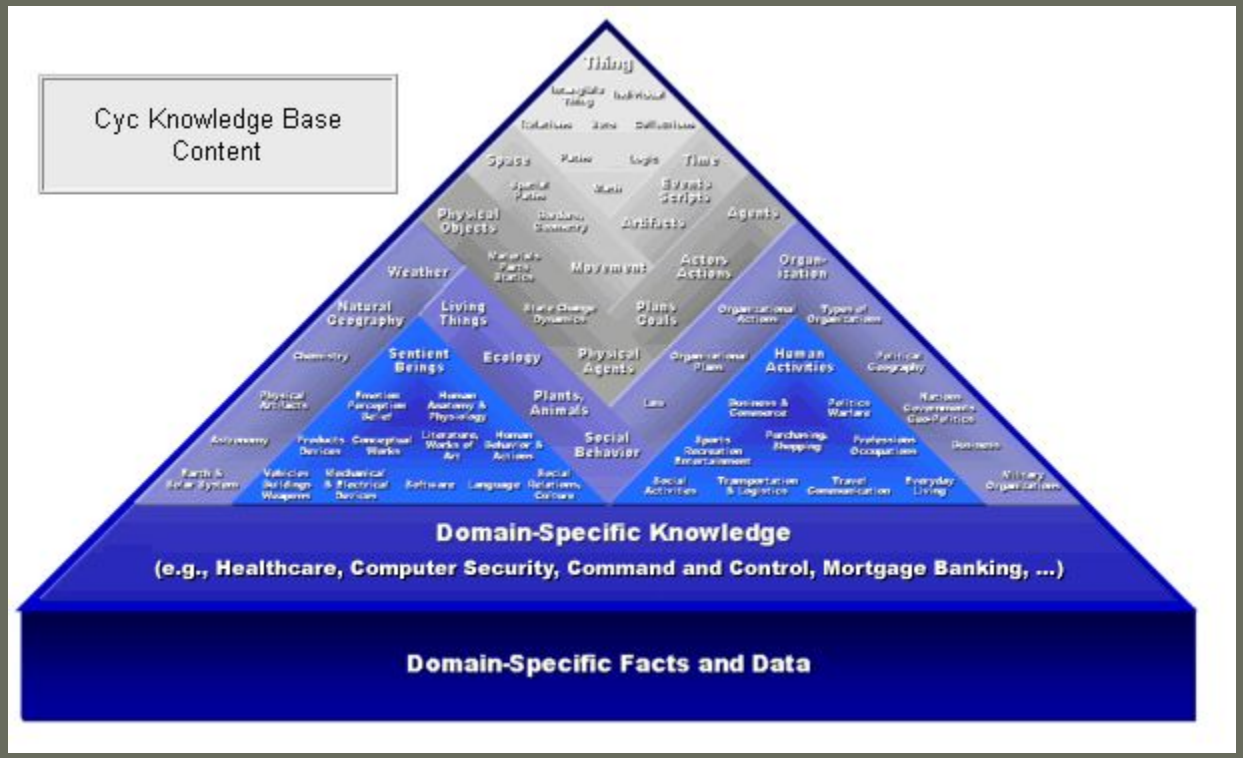
>3 500 000 утверждений

23 000 микротеорий

20 000 существительных

40 000 словосочетаний

100 000 имен собственных



Выводы

- Крупные онтологические проекты с точки зрения их функциональности не столько конкурируют, сколько дополняют возможности друг друга
- Готового полнофункционального и наполненного знаниями ресурса не существует.