**Лабораторная работа №1. Создание онтологий в Protégé**

**РАССМАТРИВАЕМЫЕ ВОПРОСЫ**

1. Формальная модель онтологии. Классификация онтологий.
2. Языки описания онтологий.
3. Средство построения онтологий Protégé.

**ФОРМАЛЬНАЯ МОДЕЛЬ ОНТОЛОГИИ**

Слово «онтология» имеет два значения:

1. Философская дисциплина, которая изучает наиболее общие характеристики бытия и сущностей;
2. Формализация предметной области для компьютерной системы.

Онтология – это подробная спецификация модели предметной области; она включает в себя словарь (т.е. список логических констант и предикатных символов) для описания предметной области и набор логических высказываний, формулирующих существующие в данной проблемной области ограничения и определяющих интерпретацию словаря. Одно из самых известных определений онтологии, сформулированное Грубером таково: Онтология – это спецификация концептуализации. Концептуализация – это структура реальности, рассматриваемая независимо от словаря предметной области и конкретной ситуации. Онтология состоит из терминов (понятий), их определений и атрибутов, а также связанных с ними аксиом и правил вывода.

Формальная модель онтологии — это четверка конечных множеств, где:

 – множество *классов*;

 – множество *атрибутов* классов;

 – множество *доменов* (областей допустимых значений) атрибутов;

 – множество *ограничений*: , где   
,  - ограничения, описывающие принадлежность атрибутов классам [17, 14];   
,  - ограничения, описывающие принадлежность доменов атрибутам;   
,  - ограничения, задающие совместимость классов (структурные ограничения совместимости классов);   
,  – ограничения, описывающие иерархические связи между классами (иерархические структурные ограничения) и включающие два типа отношений: «быть экземпляром», определяющих таксономию классов, и «быть частью», определяющих иерархию классов.

,  - ограничения, описывающие ассоциативные связи между классами (структурные ограничения одного уровня);   
, , , , ,  - функциональные ограничения, описывающие функциональные отношения между классами и атрибутами.

*Онтологии верхнего уровня* описывают общие понятия такие, как пространство, время, материя, объект, событие, действие и т.п. Онтологии верхнего уровня не зависят от конкретной проблемной области. Они определяют независимые понятия, которые могут служить именами классов (типов) при типизации объектов проблемных областей.

*Онтологии предметной области* описывают объекты, их свойства и отношения между объектами применительно к конкретной предметной области (например, медицина), используя словарь этой области.

*Онтологии задач* описывают существующие в данной проблемной области задачи (например, планирование, диагностика и т.п.). Часто задачи описываются вместе с методами их решения. В этом случае онтологии называют онтологии задач и методов. Онтологии задач включают в себя все понятия словаря проблемной области, необходимые для описания процесса логического вывода.

*Прикладные онтологии* являются специализацией онтологий предметных областей и онтологий задач. Данные онтологии задают спецификацию задач, используя словарь онтологии предметной области. Прикладные онтологии показывают, какую роль играют объекты, формализованные в онтологии предметной области, при решении той или иной задачи.

**ЯЗЫКИ ОПИСАНИЯ ОНТОЛОГИЙ**

Ключевым моментом в проектировании онтологии является выбор соответствующего языка спецификации онтологий (Ontology specification language). Цель таких языков — дать возможность указывать дополнительную машинно-интерпретируемую семантику ресурсов, сделать машинное представление данных более похожим на положение вещей в реальном мире, существенно повысить выразительные возможности концептуального моделирования слабо структурированных Web-данных.

**RDF**

Представляет собой абстрактную модель, которая определяет взаимосвязи между различными ресурсами. Модель RDF позволяет описывать предметную область с помощью базы знаний. Синтаксис использует язык XML, позволяющий обозначать факты и схемы в веб-совместимом формате и строить на их основе объектную модель для описания концептов и их взаимосвязей. В рамках проекта семантической интерпретации информационных ресурсов Интернет (Semantic Web) был предложен стандарт описания метаданных о документе Resource Description Framework, использующий XML-синтаксис. RDF использует базовую модель данных *«объект — атрибут — значение»* и способен сыграть роль универсального языка описания семантики ресурсов и взаимосвязей между ними. Для извлечения информации из RDF модели консорциумом W3C предложен специальный язык запросов и соответствующий протокол SPARQL (аббревиатура от англ. «Simple Protocol and RDF Query Language»), который позволяет искать и выполнять выборки необходимой информации из огромного массива, хранимого в World Wide Web.

**DAML+OIL** — семантический язык разметки Web-ресурсов, расширяющий стандарты RDF и RDF Schema за счет более полных примитивов моделирования. Последняя версия DAML+OIL обеспечивает богатый набор конструкций для создания онтологии и разметки информации таким образом, чтобы их могла читать и понимать машина.   
Первыми предложениями по описанию онтологии на базе RDFS были DARPA DAML-ONT (DARPA Agent Markup Language) и European Commission OIL (Ontology Inference Layer). Эти стандарты спецификации и обмена онтологиями были разработаны для поддержки процесса обмена знаниями и интеграции знаний. На базе этих предложений и возникло совместное решение DAML+OIL. Онтология DAML+OIL состоит из: заголовков (headers); элементов классов (class elements); элементов свойств (property elements); экземпляров (instances).

**OWL** (Web Ontology Language) — язык представления онтологий, расширяющий возможности XML, RDF, RDF Schema и DAML+OIL. Язык OWL предназначен для определения и представления онтологий для Семантического Веб. Язык позволяет описывать предметную область с использованием абстрактных понятий «класс» и «отношения между классами». Он расширяет схему RDFS, позволяя определять сложные взаимосвязи между различными RDFS-классами и более разнообразные ограничения на классы и их свойства. Например, можно явно ограничить количество и типы свойств класса, определить, имеют ли все экземпляры некоторого класса некоторое свойство.

**ВАРИАНТЫ ЗАДАНИЙ НА ЛАБОРАТОРНУЮ РАБОТУ**

1. Создать средствами Protégé онтологию, которая отражает структуру и состав сотрудников какой-либо организации. Полученная модель должна давать представление о том, какие подразделения входят в состав организации, каков состав сотрудников организации, какие должности занимают сотрудники, какую зарплату получают и т.п.
2. Создать средствами Protégé онтологию, которая описывает взаимодействие физического лица с визовым центром или консульством. Необходимо указать какие виды виз можно получить на въезд в выбранную страну (туристическая виза, гостевая виза, бизнес-виза), какие документы требуются для получения данной визы.