Федеральное государственное автономное образовательное учреждение высшего образования Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики



ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

ЛЕКЦИЯ 2. ДАННЫЕ, ИНФОРМАЦИЯ, ЗНАНИЯ И МУДРОСТЬ

к.т.н., Кашевник Алексей Михайлович, alexey@iias.spb.su к.т.н., Пономарев Андрей Васильевич ponomarev@iias.spb.su

Группа ВКонтакте:

https://vk.com/smartst

ПЛАН ЛЕКЦИИ



- Основные понятия, отличия данных от информации, знаний.
 Понимание.
- Модели представления знаний. Онтологии. Проектирование онтологий.
- Семантический Веб. Модель RDF. Язык запросов SPARQL.
 Модель OWL.
- Редакторы онтологий. Protege.

ПИРАМИДА: ДАННЫЕ, ИНФОРМАЦИЯ, ЗНАНИЯ, МУДРОСТЬ





« Я знаю только то, что ничего не знаю, но многие не знают даже этого» **Сократ**

ПРИМЕР







Information



Presentation



Knowledge



4

Интеллектуальные системы и технологии Лекция 2. Данные, информация, знания и мудрость

ВИДЫ ЗНАНИЙ



Предметные знания – количественные и качественные характеристики конкретных объектов, явлений и их элементов на выбранном представлении.

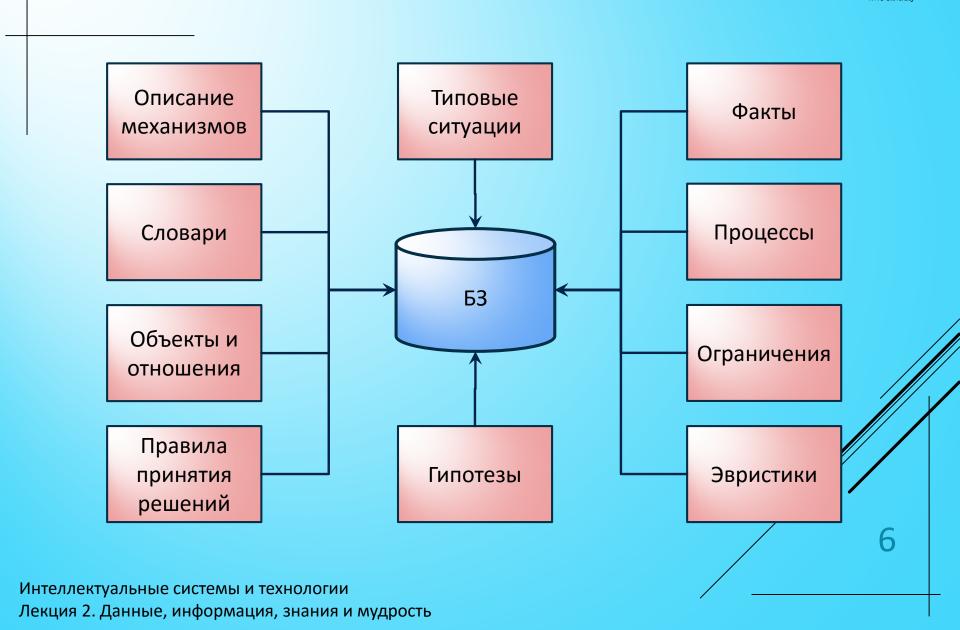
Процедурные знания – методы, алгоритмы и программы выполнения различных действий.

Понятийные знания — совокупность понятий, терминов, свойств, взаимосвязей понятий, используемых в некоторой предметной области.

Конструктивные знания — знания о возможной структуре и взаимодействии частей различных объектов предметной области.

ТИПЫ ЗНАНИЙ





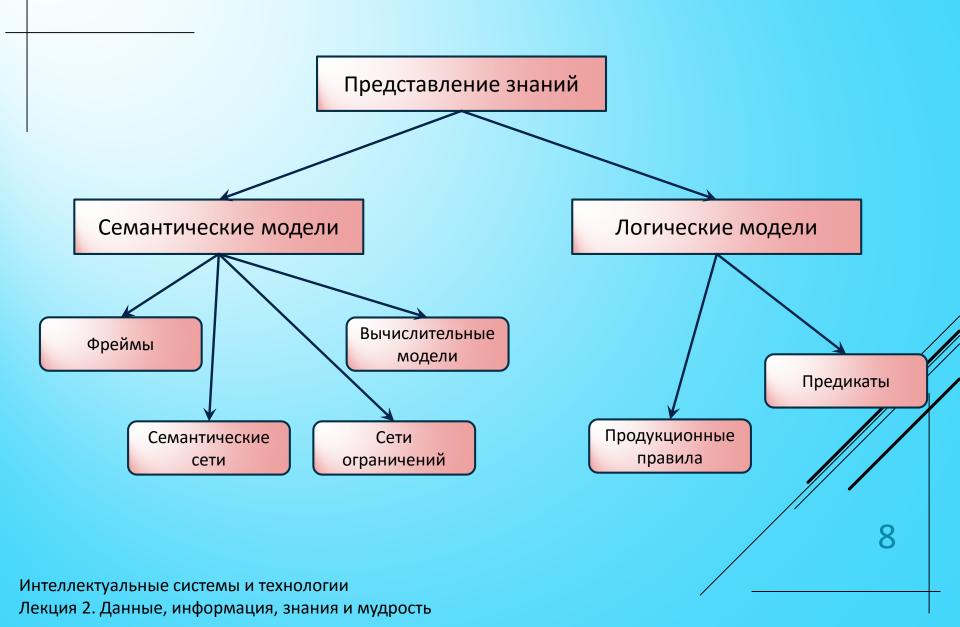
ПРИНЦИПЫ ПОСТРОЕНИЯ МОДЕЛИ ПРЕДМЕТНОЙ ОБЛАСТИ



- Адекватность полнота отражения объектов и связей
- <u>Открытость</u> возможность дополнения модели знаниями
- Прослеживаемость ассоциативных связей между описаниями объектов и ситуаций
- <u>Иерархичность</u> поиск информации об определённом объекте не должен сопровождаться просмотром всех объектов модели, а должны рассматриваться только объекты связанные с данным

МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ

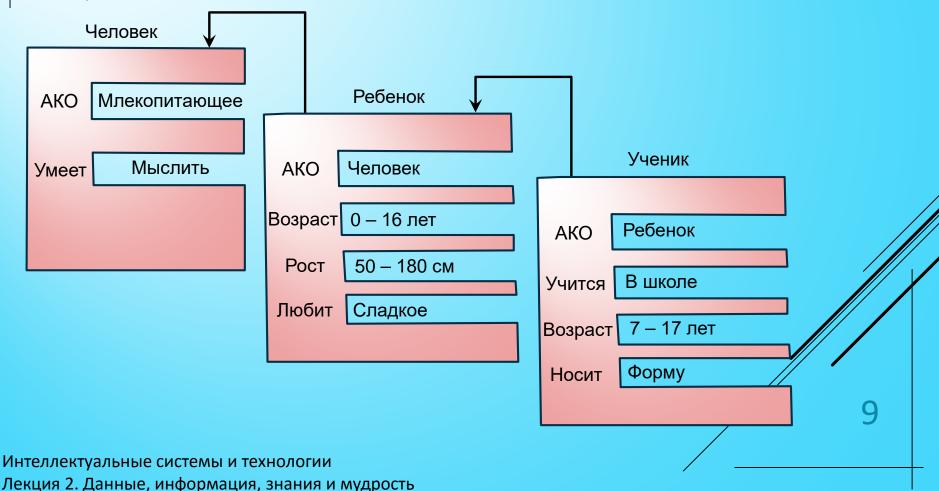




МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ: ФРЕЙМЫ



Фреймы представляют собой структуры для описания пространственных сцен, типовых объектов, событий, стереотипных ситуаций.



ПРЕИМУЩЕСТВА И НЕДОСТАТКИ ФРЕЙМОВЫХ МОДЕЛЕЙ



Достоинства:

- + Отражает концептуальную основу организации памяти человека [Шенк, Хантер, 1987].
- + Позволяют обеспечить механизм наследования
- + Позволяют формально связывать модульный принцип организации прикладных программных средств с информационными структурами.

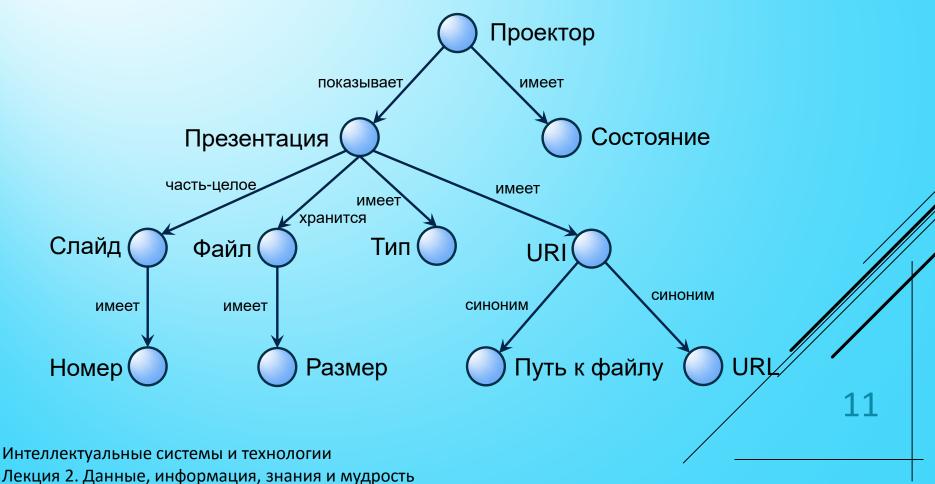
Недостатки:

- Отсутствует возможность интерпретации отношений произвольного типа (только таксономические отношения).
- Механизмы вывода для фреймового представления слабо развиты.

МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ: СЕМАНТИЧЕСКИЕ СЕТИ



Семантическая сеть – это ориентированный граф, вершины которого – понятия, а дуги – отношения между ними



Лекция 2. Данные, информация, знания и мудрость

ВИДЫ СЕМАНТИЧЕСКИХ СЕТЕЙ



По количеству типов отношений

- Однородные сети с единственным типом отношений.
- Неоднородные сети с различными типами отношений.

По типам отношений

- Бинарные сети, в которых отношения связывают два объекта.
- **N-арные** сети, в которых есть специальные отношения, связывающие более двух понятий.

Обычно в семантических сетях используются следующие отношения:

- связи типа: «часть целое» («класс подкласс», «элемент множество»);
- функциональные связи;
- количественные связи (больше, меньше, равно и т. п.);
- пространственные (далеко от, близко от, и т. п.);
- временные (**раньше**, **позже** и т. п.);
- атрибутивные связи (иметь свойство, иметь значение);
- логические связи (и, или, не);
- Лингвистические связи;
- и др.

ПРЕИМУЩЕСТВА И НЕДОСТАТКИ СЕМАНТИЧЕСКИХ СЕТЕЙ



Достоинства:

- + Направленность на решение проблемы информационного поиска.
- + Удобная графическая нотация.
- + Универсальность, достигаемая за счет выбора соответствующего набора отношений, позволяющая описать сколь угодно сложную ситуацию, факт или предметную область.

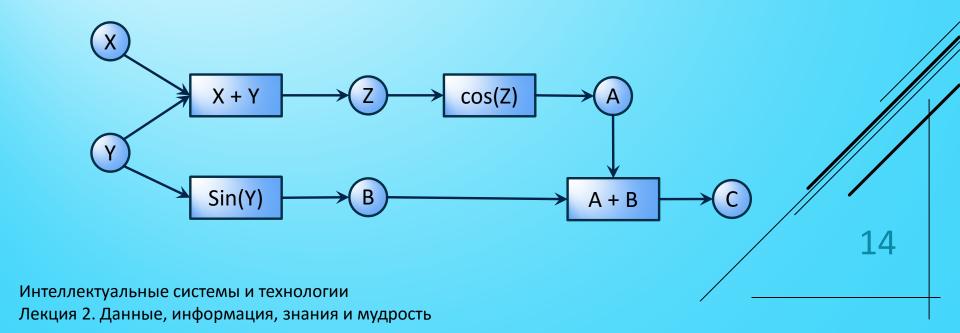
Недостатки:

- Сетевые модели представляют собой пассивные структуры, для обработки которых необходим специальный аппарат формального вывода.
- Представление, использование и модификация знаний при описании систем реального уровня сложности оказывается трудоемкой процедурой, особенно при наличии множественных отношений между ее понятиями.

МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ: ВЫЧИСЛИТЕЛЬНЫЕ МОДЕЛИ



Вычислительная модель, или вычислительная (функциональная) сеть представляет собой двудольный ориентированный граф, включающий вершины двух типов — объекты и операторы (функции). Дуги отражают функциональные связи между операторами и объектами. Дуга, направленная от объекта к оператору, предписывает рассматривать этот объект как аргумент данного оператора, дуга обратной ориентации указывает на то, что объект выступает по отношению к оператору в качестве результата.



ДОСТОИНСТВА И НЕДОСТАТКИ ФУНКЦИОНАЛЬНЫХ МОДЕЛЕЙ

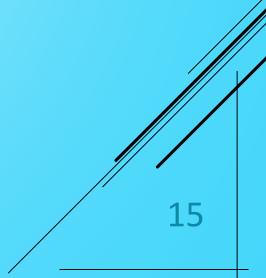


Достоинства:

- Народний от процедуры вычислений одних объектов через другие.
- + Функциональная сеть имеет наглядное представления.

Недостатки:

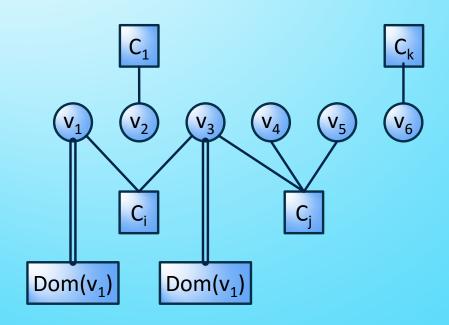
- Сложность описания комплексных задач.
- Может использоваться для узкого круга задач.



МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ: СЕТИ ОГРАНИЧЕНИЙ



Сеть ограничений: (V, dom, C) включает в себя множество переменных $V=\{v_1, ..., v_i, ..., v_n\}$, каждая из которых принимает значение в соответствующей области dom(i), dom(j), ..., dom(n) и множества ограничений С. Каждое ограничение $c(i_1, ..., i_q)$, ограничивающее подмножество $(i_1, ..., i_q)$ переменных v определяет, какие из значений переменных согласуются друг с другом.

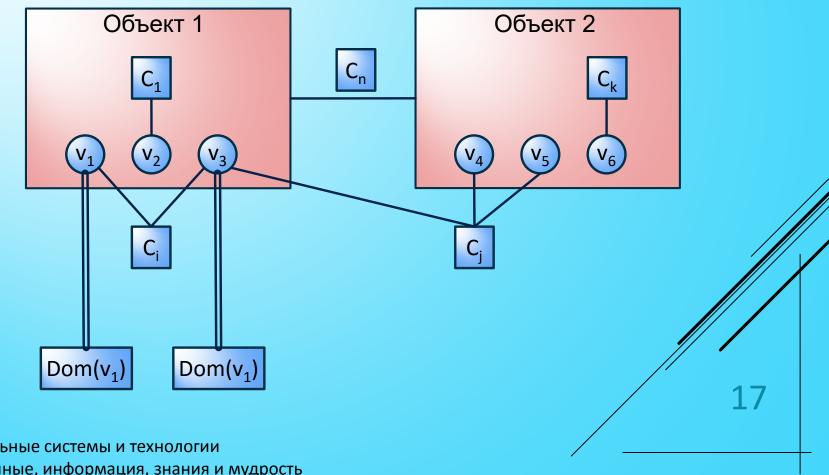




МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ: ОБЪЕКТНО-ОРИЕНТИРОВАННЫЕ СЕТИ ОГРАНИЧЕНИЙ



Позволяют объединить несколько переменных в объект и оперировать терминами ограничения между объектами



Интеллектуальные системы и технологии Лекция 2. Данные, информация, знания и мудрость

ЛОГИЧЕСКИЕ МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ: ПРОДУКЦИОННЫЕ ПРАВИЛА



Продукционная модель или модель основанная на правилах позволяет представить знания в виде предложений типа:

<имя правила> ЕСЛИ <условие> TO <действие>

Под условием понимается некоторое предложение-образец, по которому осуществляется поиск в базе знаний, а под действием понимается то, что выполняется при успешном исходе поиска.

Продукционная модель: $i = \langle S; L, A \rightarrow B; Q \rangle$, где

S – описание класса ситуаций;

L – условие, при котором продукция активизируется;

 $A \rightarrow B$ – ядро продукции;

Q – постусловие продукционного правила.

ЕСЛИ

«двигатель не заводится» **и** «стартер двигателя не работает» **то**

«неполадки в системе электропитания стартера»

ЛОГИЧЕСКИЕ МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ: ПРЕДИКАТЫ



Информация, необходимая для решения прикладных задач, рассматривается как совокупность фактов и утверждений, которые представляются как формулы в некоторой логике. Знания отображаются совокупностью таких формул, а получение новых знаний сводится к реализации процедур логического вывода.

Предикатом называется функция, принимающая два значения (истина или ложь) и предназначенная для выражения свойств объектов или связей между ними.

Выражение, в котором утверждается или отрицается наличие каких-либо свойств у объекта, называется высказыванием.

ДАТЬ (МИХАИЛ, ВЛАДИМИРУ, КНИГУ);

ПРЕИМУЩЕСТВА И НЕДОСТАТКИ ЛОГИЧЕСКИХ МОДЕЛЕЙ



Достоинства:

- + Хорошо изученный и обоснованный классический аппарат математической логики.
- + Имеют ясную формальную семантику, существование достаточно эффективных процедур вывода.
- + В базах знаний можно хранить лишь множество аксиом, а все остальные знания получать из них по правилам вывода.

Недостатки:

- С помощью правил, задающих синтаксис языка, нельзя установить истинность или ложность того или иного высказывания, высказывание может быть построено синтаксически правильно, но оказаться совершенно бессмысленным.
- отсутствие средств для структурирования используемых элементов и недопустимость противоречий

УПРАВЛЕНИЕ ЗНАНИЯМИ



Управление знаниями (менеджемент знаний) - это дисциплина, которая обеспечивает интегрированный подход к созданию, сбору, организации и использованию

информационных ресурсов и доступу к ним. Эти ресурсы включают структурированные БД, текстовую информацию, такую, как документы, и, что наиболее важно, неявные знания и накопленный опыт сотрудников.

ОНТОЛОГИИ



Онтология — это подробная спецификация модели предметной области; она включает в себя <u>словарь</u> (т.е. список логических констант и предикатных символов) для описания предметной области и <u>набор логических высказываний</u>, формулирующих существующие в данной проблемной области ограничения и определяющих интерпретацию словаря.

НОТАЦИЯ ОБЪЕКТНО-ОРИЕНТИРОВАННЫХ СЕТЕЙ ОГРАНИЧЕНИЙ ДЛЯ ПРЕДСТАВЛЕНИЯ ОНТОЛОГИЙ

- ightharpoonup O = < C, A, D, R >, где
 - ightharpoonup C множество классов;
 - ► A множество атрибутов классов;
 - ▶ D множество доменов атрибутов;
 - ► R множество ограничений.
- Ограничения описывают отношения между элементами онтологии, а также функциональные зависимости между значениями атрибутов классов:
 - Принадлежность атрибутов классам;
 - ▶ Принадлежность доменов атрибутам;
 - ▶ Совместимость классов;
 - ▶ Иерархические связи («быть экземпляром» и «быть частьум»);
 - Ассоциативные связи;
 - Функциональные связи

ПРОЕКТИРОВАНИЕ ОНТОЛОГИЙ: ФУНДАМЕНТАЛЬНЫЕ ПРАВИЛА



- 1) Не существует одной правильной модели предметной области, всегда возможны альтернативные варианты построения онтологии. Лучшее решение всегда зависит от существующего приложения и ожидаемых дополнений.
- 2) Построение онтологий это обязательно итеративный процесс.
- 3) Концепты онтологии должны быть близки к объектам и их взаимосвязям рассматриваемой предметной области. Это обычно существительные для объектов и глаголы для взаимосвязей из предложений, описывающих предметную область.

ПРОЕКТИРОВАНИЕ ОНТОЛОГИЙ: АЛГОРИТМ

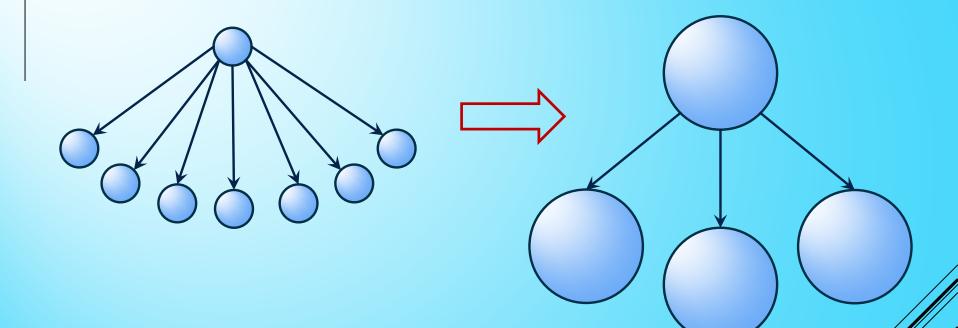


Алгоритм онтологического инжиниринга «для чайников» (проф. Т. А. Гаврилова):

- выделение концептов базовых понятий данной предметной области;
- определение «высоты дерева онтологий» числа уровней абстракции;
- распределение концептов по уровням;
- построение связей между концептами определение отношений и взаимодействий базовых понятий;
- консультации с различными специалистами для исключения противоречий и неточностей.

ПРИНЦИП ОККАМА





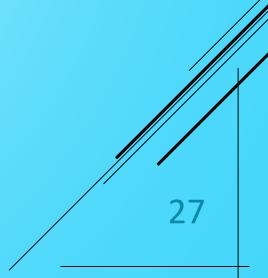
Не следует привлекать новые сущности без крайней на то необходимости

ШАБЛОНЫ ПРОЕКТИРОВАНИЯ ОНТОЛОГИЙ



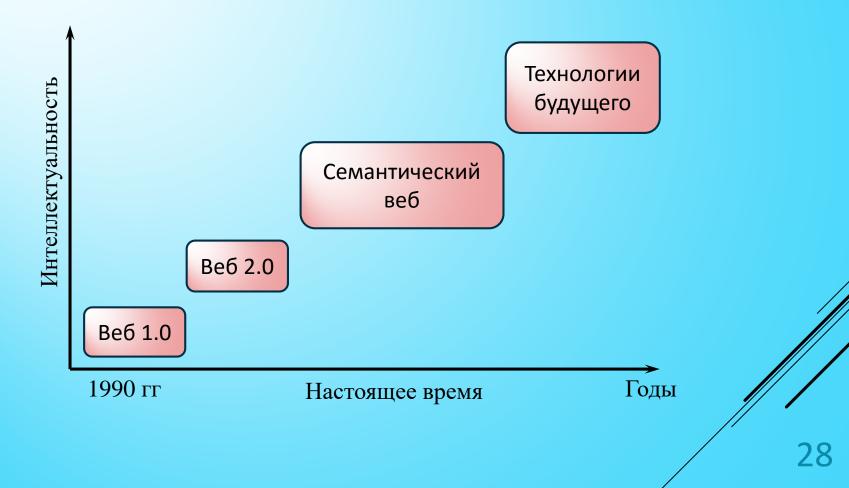
- Простое и эффективное решение конкретной типичной задачи
- Быстрый обмен опытом между разработчиками
- Массовость внедрения

http://ontologydesignpatterns.org/



СЕМАНТИЧЕСКИЙ ВЕБ ЭВОЛЮЦИЯ





Интеллектуальные системы и технологии Лекция 2. Данные, информация, знания и мудрость

СЕМАНТИЧЕСКИЙ ВЕБ



- Модели **RDF(S)** и **OWL** для представления знаний предметной области.
- XML (eXtensible Markup Language), определяющая синтаксис и структуру инф. В сети Интернет
- **RDF** (Resource Description Framework) средство описания ресурсов Семантического Веб.
- **OWL** (Ontology Web Language) язык описания онтологий, базирующийся на RDF.
- Объединяет в себе семантические и логические модели представления знаний.
- Protégé самый известный на сегодняшний день редактор онтологий, поддерживающий RDF и OWL.

RDFS (RDF SCHEMA)



RDFS – язык описания словарей для RDF.

RDF Schema определяет классы, свойства и другие ресурсы.

RDFS является семантическим расширением RDF.

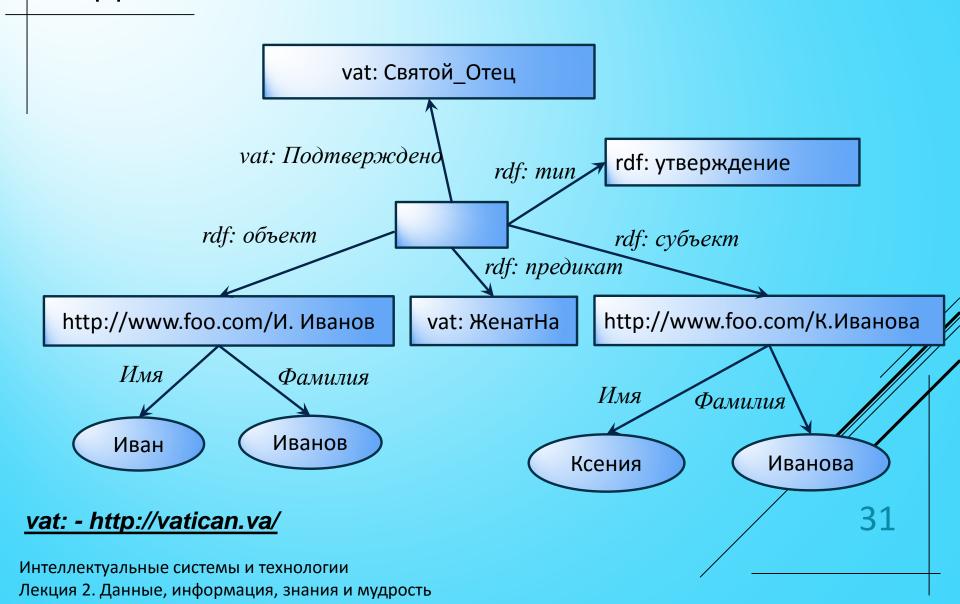
Пример:

http://vatican.va/Святой_Отец



ГРАФИЧЕСКИЙ ПРИМЕР RDF(S) МОДЕЛИ





ЯЗЫК SPARQL ДЛЯ ПОИСКА ИНФОРМАЦИИ В RDF



```
PREFIX foo: <a href="http://example.com/resources/">http://example.com/resources/</a>
# префиксные объявления
FROM ...
# источники запроса
SELECT ...
# состав результата
WHERE {...}
# шаблон запроса
ORDER BY ...
# модификаторы запроса
Пример:
SELECT ?person
WHERE
 <a href="http://www.foo.com/И.Иванов"> <ЖенатНа> ?person</a>
```

МОДЕЛЬ OWL



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.

- OWL Lite: позволяет, описывать таксономию предметной области и простые ограничения. По сравнению с другими разновидностями OWL, поддержка OWL Lite проще для программной реализации.
- **OWL DL** (аббревиатура DL от англ. descriptive logic): предназначена для пользователей, которым необходима максимальная выразительность (в терминах дескрипционной логики) при сохранении полноты вычислений и разрешимости. Все логические заключения, подразумеваемые той или иной онтологией, будут вычислимыми. В вычисления завершатся с корректным результатом за конечное время.
- **OWL Full**: предназначена для использования максимальной выразительности и синтаксической свободы модели <u>RDF</u>. При этом нетрантий вычислимости.

Интеллектуальные системы и технологии Лекция 2. Данные, информация, знания и мудрость

СТРУКТУРА OWL ОНТОЛОГИИ



- Заголовок
 - версия
 - примечания
 - импортируемые онтологии
- Тело
 - описания классов, свойств и индивидов в форме аксиом

OWL. БАЗОВЫЕ ЭЛЕМЕНТЫ. КЛАССЫ





<owl:Class rdf:ID="Human"/> - определение именованного класса с именем "Human"

rdfs:subClassOf - говорит о том, что один класс полностью входит в другой.

owl:equivalentClass - говорит о том, что два класса совпадают.

owl:disjointWith - говорит о том, что два класса не пересекаются.

Пример: SavingAccount - это класс, являющийся подклассом класса Account.

<owl:Class rdf:ID="SavingsAccount">

<rdfs:subclassOf rdf:resource="#Account"/>

</owl:Class>

OWL. БАЗОВЫЕ ЭЛЕМЕНТЫ. СВОЙСТВА



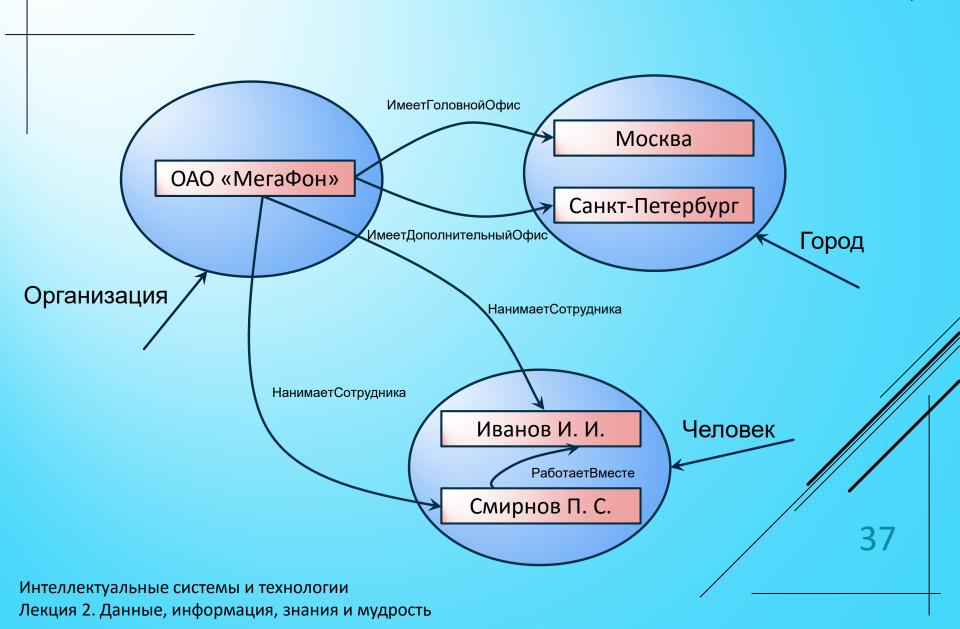
Индивиды - это элементы классов; свойства могут связывать их друг с другом. Например, индивид Smith может быть описан как элемент, принадлежащий классу Person.

owl:ObjectProperty - объектные свойства, связывающие между собой индивиды

owl:DatatypeProperty - свойства-значения, связывающие индивиды со значениями данных

ПРИМЕР OWL ОНТОЛОГИИ





РЕДАКТОРЫ ОНТОЛОГИЙ



Protégé (http://protege.stanford.edu/)

WebODE (http://webode.dia.fi.upm.es/webODE/)

OntoSaurus (http://www.isi.edu/isd/ontosaurus.html)

WebOnto (http://kmi.open.ac.uk/projects/webonto/)

OilEd (http://oiled.man.ac.uk/)

OntoEdit (http://ontoserver.aifb.unikarlsruhe.de/ontoedit/)

OilEd (http://oiled.man.ac.uk/)

Ontolingua (http://www-ksl.stanford.edu)

PEДАКТОР PROTÉGÉ



- Изначально создавался для концептуального моделирования в области медицины
- В последнее время используется при создании онтологий для Семантического Веба
- Поддерживаемые языки представления онтологий: DAML+OIL, RDF/RDFS, OWL

ОБЩЕДОСТУПНЫЕ БИБЛИОТЕКИ ОНТОЛОГИЙ



- DAML ontology library (http://www.daml.org/ontologies/)
- Protege ontology library (http://protege.stanford.edu/ontologies.html)
- Ontolingua ontology library (http://ontolingua.stanford.edu/)
- WebOnto ontology library (http://webonto.open.ac.uk)
- SHOE ontology library (http://www.cs.umd.edu/projects/plus/SHOE/onts/index.html)
- WebODE ontology library (http://webode.dia.fi.upm.es/)
- (KA)2 ontology library (http://ka2portal.aifb.uni-karlsruhe.de/)
- AKT ontology (http://www.aktors.org/ontology/)

СЕГОДНЯ МЫ УЗНАЛИ

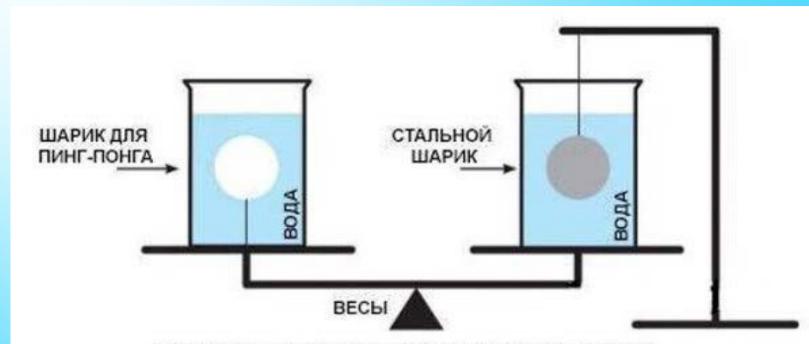


- 1) Понятия данные, информация, знания и их отличия
- 2) Основные модели представления знаний
- 3) Понятие онтологии, проектирование онтологий
- 4) Семантический Веб
- 5) Модель RDF и Язык запросов SPARQL
- 6) Moдeль OWL
- 7) Редакторы онтологий
- 8) Редактор Protege



ЗАДАЧА ПРО КОЛБЫ





ДИАМЕТРЫ ШАРОВ РАВНЫ. ОБЪЁМ ВОДЫ ОДИНАКОВ. ДАВЛЕНИЕ АТМОСФЕРНОЕ. ОБЪЁМОМ НИТИ ПРЕНЕБРЕЧЬ.

НАЛЕВО ИЛИ НАПРАВО?