ЛЕКЦИЯ №8





ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ: ПРЕДСТАВЛЕНИЕ ЗНАНИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

- Данные и знания.
- Модели представления знаний.
- Технологии баз знаний в Интернете.

что же такое знания ???

- чем они отличаются от обычных данных, десятилетиями обрабатываемых компьютерами?

Данные – это представление фактов и идей в формализованном виде, пригодном для передачи и обработки в некотором информационном процессе.

Знания – итог теоретической и практической деятельности человека, отражающий накопление предыдущего опыта и отличающийся высокой степенью структурированности, правила использования этой информации для принятия решений.



Главное отличие знаний от данных состоит в их активности, то есть появление новых фактов или установление новых связей может стать источником активности системы.

Знания



ДЕКЛАРАТИВНЫЕ ЗНАНИЯ

Я знаю, **что** . . .

Знания:

- *о явлениях*
- *о событиях*
- *о свойствах объектов*
- *о зависимостях*

ПРОЦЕДУРНЫЕ ЗНАНИЯ

Я знаю, **как** . . .

Знания, определяющие действия для достижения какой-либо цели

База знаний

Для хранения данных используются базы данных (БД), для которых характерны большой объем и относительно небольшая удельная стоимость информации. Для хранения знаний соответственно применяются базы знаний (БЗ), обладающие зачастую небольшим объемом, но являющиеся исключительно дорогими информационными массивами.

- **База знаний** – основа любой интеллектуальной системы. Раздел искусственного интеллекта, изучающий базы знаний и методы работы со знаниями, называется *инженерией знаний*.



Классы МПЗ

Существуют десятки моделей представления знаний для различных предметных областей. Большинство из них может быть сведено к следующим классам:

- продукционные модели;
- семантические сети;
- фреймы;
- формальные логические модели



Модели представления знаний



Продукционная модель

- **Продукционная модель** (модель, основанная на правилах) позволяет представить знания в виде предложений, называемых продукциями, типа «**Если** (условие), **то** (действие)».
- Под условием (антецедентом) понимается некоторое предложение-образец, по которому осуществляется поиск в БЗ, а под «действием» (консеквентом) – операции, выполняемые при успешном исходе поиска.
- Существует большое количество программных средств, реализующих продукционный подход: язык OPS 5, оболочки ЭС – EXSYS Professional, Kappa, ЭКСПЕРТ, инструментальные системы ПИЭС и СПЭИС и др.



Недостаток: при накоплении достаточно большого количества (порядка нескольких сотен) продукций они начинают противоречить друг другу.

Семантические сети

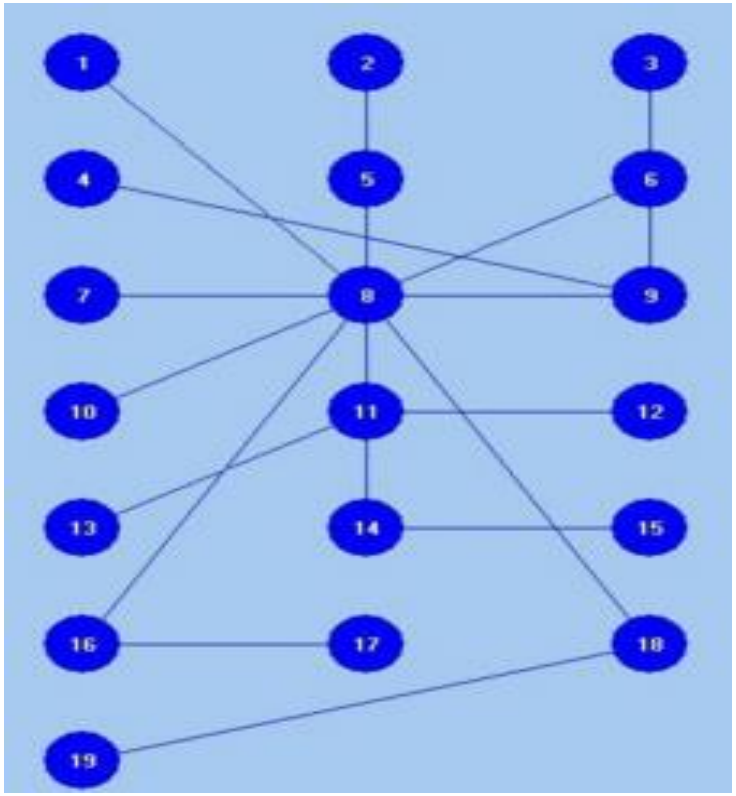
Семантическая сеть – это ориентированный граф, вершины которого отображают некоторые понятия, а дуги – отношения между ними (отражает семантику предметной области в виде понятий и отношений).

- Наиболее часто возникает потребность в описании отношений между элементами, множествами и частями объектов.
- Отношение между *объектом* и *множеством*, обозначающее, что объект принадлежит этому множеству, называется отношением классификации (**ISA**). Связь ISA предполагает, что свойства объекта наследуются от множества.
- Обратное к ISA отношение используется для обозначения примером, поэтому так и называется – «**Example**».
- Отношение между надмножеством и подмножеством называется **AKO** (A Kind Of).

отношения

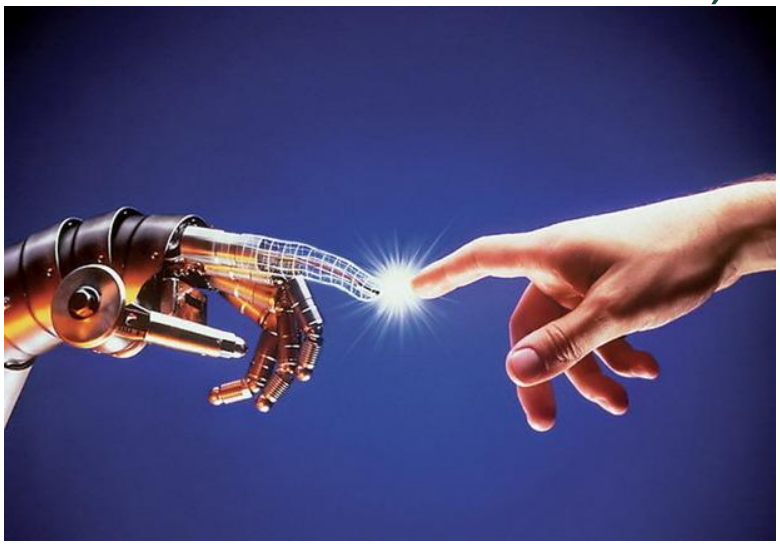
- Элемент подмножества называется **гипонимом**, а надмножества – **гиперонимом**, само же отношение называется **отношением гипонимии**.

- **Оношение гипонимии** определяет, что каждый элемент первого множества входит и во второе (выполняется ISA для каждого элемента), а также логическую связь между самими подмножествами: что первое не больше второго и свойства первого множества наследуются вторым.



Используются также следующие отношения:

- функциональные связи (определяемые обычно глаголами
- «производит», «влияет» и др.);
- количественные (больше, меньше, равно);
- пространственные (далеко от, близко к, за, под, над);
- временные (раньше, позже, в течение
- атрибутивные (иметь свойство, иметь значение);
- логические (и, или, не);
- лингвистические.



Языки семантических сетей

- Для реализации семантических сетей существуют специальные сетевые языки, например NET, язык реализации систем SIMER+MIR и др.
- Широко известны экспертные системы, использующие семантические сети в качестве языка представления знаний – PROSPECTOR, CASNET, TORUS.



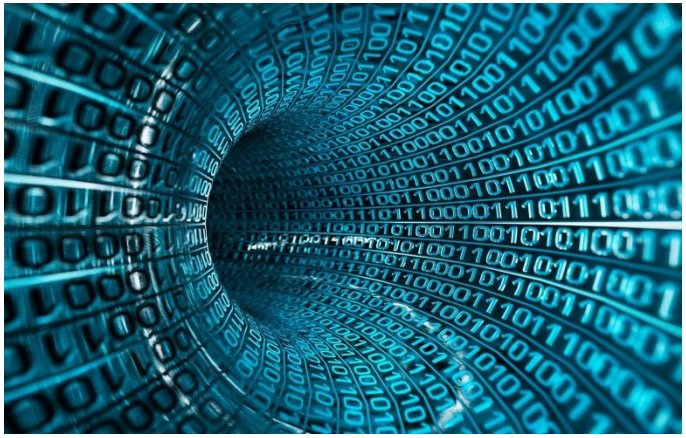


Фреймы

Фрейм – это абстрактный образ для представления некоего стереотипа информации.

- Фреймом также называется и формализованная модель для отображения образа. Различают фреймы-образцы (прототипы), хранящиеся в базе знаний, и фреймы-экземпляры, которые создаются для отображения реальных фактических ситуаций на основе поступающих данных.

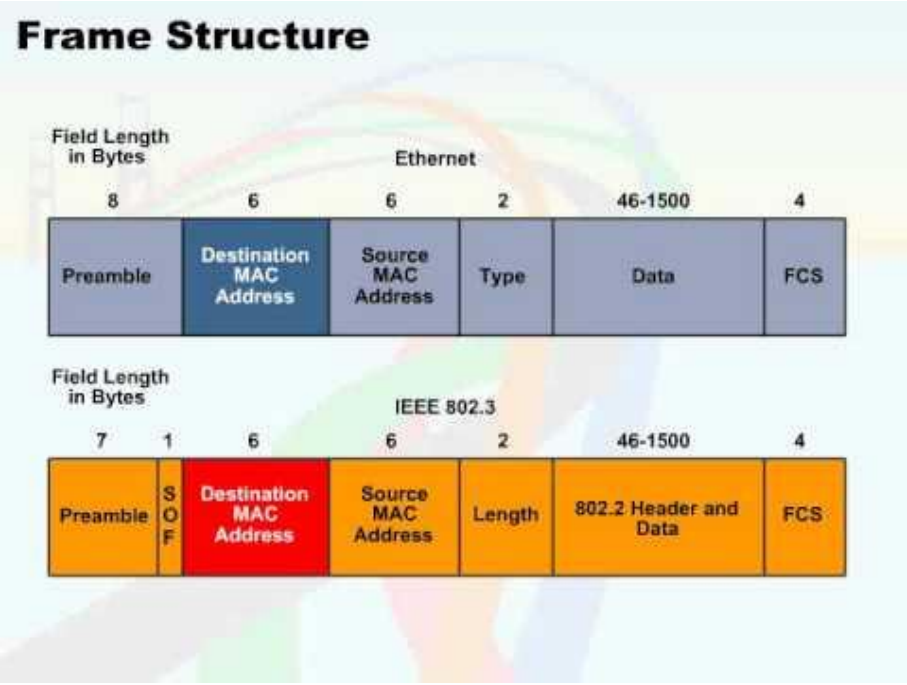




структура фрейма

Традиционно структура фрейма может быть представлена как список свойств:

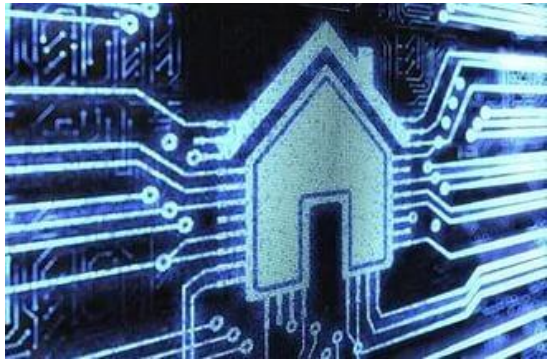
- (ИМЯ ФРЕЙМА (имя 1-го слота: значение 1-го слота), (имя 2-го слота: значение 2-го слота)
- (имя N-го слота: значение N-го слота)).



Способы получения слотом значений

В качестве значения слота может выступать имя другого фрейма: так образуются сети фреймов. Существует несколько способов получения слотом значений во фрейме-экземпляре:

- по умолчанию от фрейма-образца;
- через наследование свойств от фрейма, указанного в слоте AKO;
- по формуле, указанной в слоте;
- через присоединенную процедуру;
- явно из диалога с пользователем;
- из базы данных.



наследование свойств

- Важнейшим свойством теории фреймов является заимствование из теории семантических сетей – так называемое наследование свойств. И во фреймах, и в семантических сетях наследование происходит по AKO-связям. Слот AKO указывает на фрейм более высокого уровня иерархии, откуда неявно наследуются, т.е. переносятся, значения аналогичных слотов.
- Основным преимуществом фреймов как модели представления знаний является то, что она отражает концептуальную основу организации памяти человека, а также ее гибкость и наглядность.
- Специальные языки представления знаний в сетях фреймов FRL (Frame Representation Language), KRL (Knowledge Representation Language), фреймовая оболочка Kappa и другие программные средства позволяют эффективно строить промышленные ЭС.

Формальные логические модели

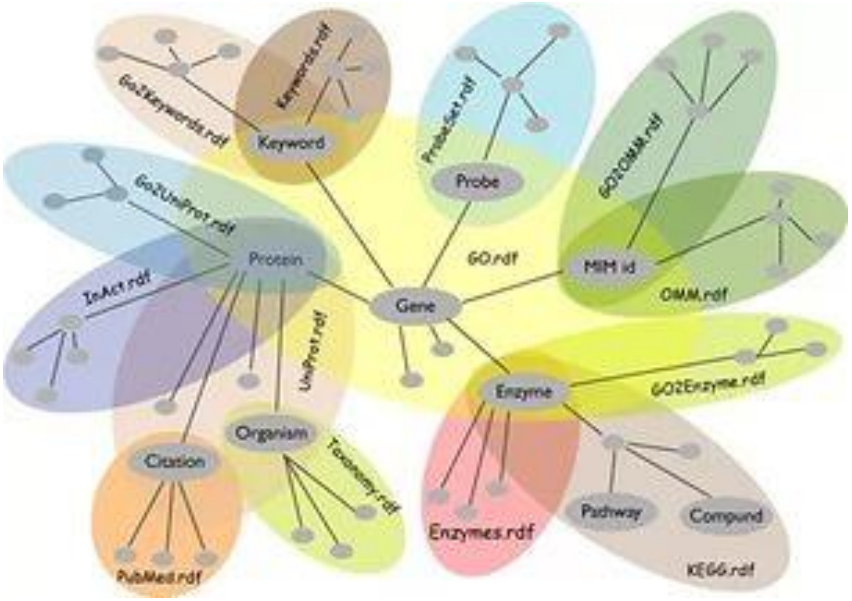
- **Логические модели** строятся при помощи декларативных языков логического программирования, наиболее известным представителем которых является язык ***Пролог*** (Prolog).

Базовым принципом языка является равнозначность представления программы и данных (декларативность), отчего утверждения языка одновременно являются и записями, подобными записям в базах данных, и правилами, несущими в себе способы их обработки. Сочетание этих качеств приводит к тому, что по мере работы системы Пролога знания (и данные, и правила) накапливаются. Поэтому Пролог-системы считают естественной средой для накопления базы знаний.



Онтологии

- **Онтология** – это попытка всеобъемлющей и детальной формализации некоторой области знаний с помощью концептуальной схемы.
- Современные онтологии строятся по большей части одинаково, независимо от языка написания. Обычно они состоят из экземпляров, понятий, атрибутов и отношений.

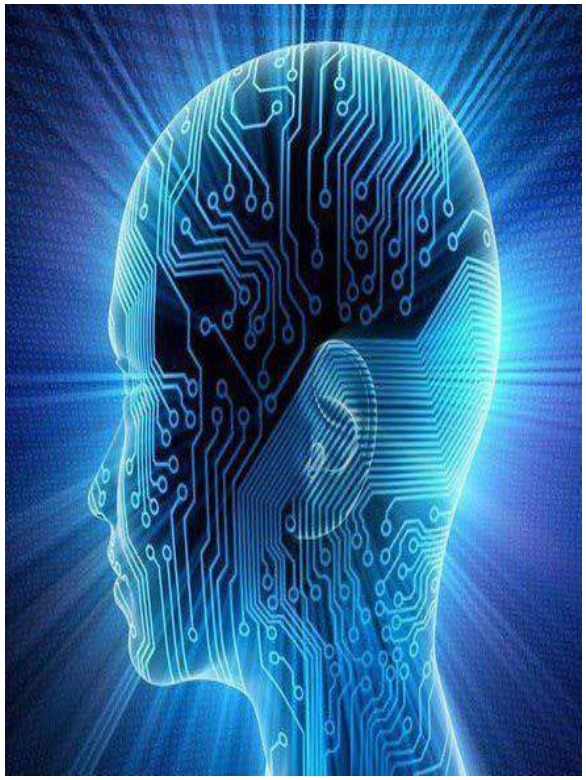


Онтологии

- **Экземпляры** (или индивиды) – это основные, нижнеуровневые компоненты онтологии. Экземпляры могут представлять собой как физические объекты (люди, дома, планеты), так и абстрактные (числа, слова).
- **Понятия** (или классы) – это абстрактные группы, коллекции или наборы объектов. Они могут включать в себя экземпляры, другие классы, либо же сочетания того и другого.
- **Объекты** в онтологии могут иметь атрибуты. Каждый атрибут имеет по крайней мере имя и значение, и используется для хранения информации, которая специфична для объекта и привязана к нему.

Онтологии

- **Специализированные** (предметно-ориентированные) **онтологии** – это представление какой-либо области знаний или части реального мира.



- Такие онтологии содержат базовый набор терминов, глоссарий или тезаурус, используемый для описания терминов предметных областей.
- Если использующая специализированные онтологии система развивается, то может потребоваться их объединение, и для инженера по онтологиям это серьезная задача

Языки описания онтологий

Разработано несколько формальных языков для описания онтологий, в частности, следующие:

- OWL (Ontology Web Language), язык для поддержки семантической паутины (см. ниже);
- KIF (Knowledge Interchange Format) – основанный на т.н. S-выражениях синтаксис для логики;
- CycL — онтологический язык, используемый в проекте Cyc, основан на исчислении предикатов с некоторыми расширениями более высокого порядка.



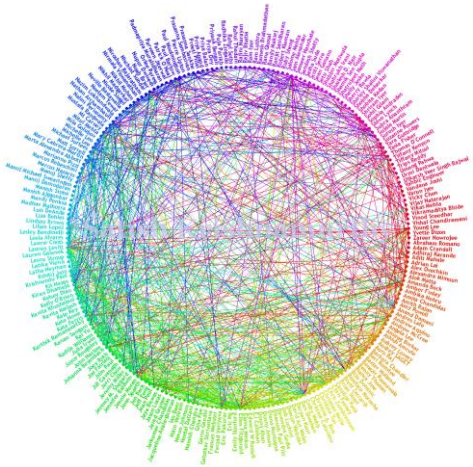


Семантическая паутина

- **Семантическая паутина -** часть глобальной концепции развития сети Интернет, целью которой является реализация возможности машинной обработки информации, доступной во Всемирной паутине.



Основной акцент концепции делается на работе с метаданными, однозначно характеризующими свойства и содержание ресурсов Всемирной паутины, вместо используемого в настоящее время текстового анализа документов.



Семантическая паутина

- Термин «семантическая паутина» впервые введен сэром Тимом Бернерсом-Ли в мае 2001 года в журнале Scientific American, и называется им «следующим шагом в развитии Всемирной паутины».
- В семантической паутине предполагается повсеместное использование, во-первых, универсальных идентификаторов ресурсов (URI), а во-вторых, — онтологий и языков описания метаданных.
- Техническую часть семантической паутины составляет семейство стандартов на языки описания, включающее XML, XML Schema, RDF, RDF Schema, OWL, а также некоторые другие.
- Форматы описания метаданных в семантической паутине предполагают проведение логического вывода на этих метаданных, и разрабатывались с оглядкой на существующие математические формализмы в этой области.





Спасибо за внимание