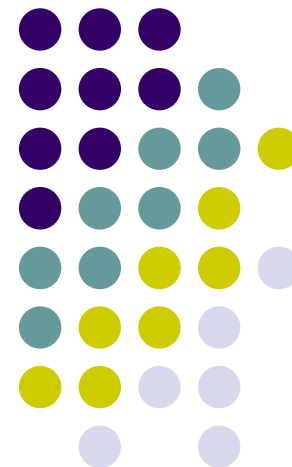


# Онтологический подход и его использование для структурирования предметных областей

Научная библиотека им. Н.И. Лобачевского Казанского  
государственного университета  
Кузьмина Варвара Юрьевна, заведующая отделом автоматизации





# Что такое онтология ?

- На протяжении веков философы пытались найти универсальные категории для классификации всего, что существует;
- лексикографы стремились к универсальной терминологии для определения всего, что можно сказать;
- библиотекари стремятся к универсальным системам предметизации и классификации для хранения и поиска всего, что было написано;
- в 70-х годах 20 века ANSI SPARC предложил три схемы архитектуры для разработки и интеграции баз данных, которые управляют мировой экономикой;
- семантический веб расширил задачу до уровня классификации, маркировки, определения, поиска, интеграции, и использования всего World Wide Web, который быстро становится универсальным хранилищем всех накопленных знаний, информации, данных и мусора человечества.

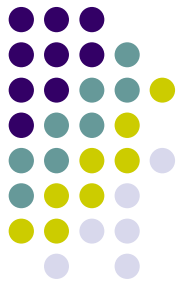


# Что такое онтология ?

**Онтология – точная (выраженная формальными средствами) спецификация концептуализации. (Gruber)**

- **Концептуализация** процесс перехода от представления предметной области на естественном или ограниченном естественном языке к точной спецификации этого описания на некотором формальном языке, ориентированном на компьютерное представление. Концептуализация также трактуется как результат подобного процесса, т. е. описание множества понятий (концептов) предметной области, знаний о них и связях (отношениях) между ними.
- **Концептуализация** – структура реальности, независимо от
  - Словаря
  - Конкретной ситуации
- **Концептуализация** - процедура введения онтологических представлений в накопленный массив эмпирических данных;
- первичная теоретическая форма, обеспечивающая теоретическую организацию материала;
- схема связи понятий и т.д.

# Интерпретации понятия «онтология»



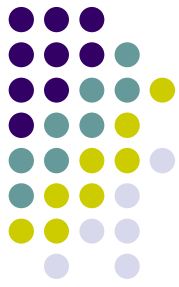
1. неформальная концептуальная система (представление концептуализации);
2. формальный взгляд на семантику;
3. спецификация концептуализации;
4. представление концептуальной системы через логическую теорию;
5. словарь, используемый логической теорией;
6. метауровневая спецификация логической теории.



# Что такое онтология ?

- Онтологией называют:
  - словарь с определениями;
  - простую таксономию;
  - тезаурус (таксономия с терминами);
  - модель с произвольным набором отношений;
  - таксономию и произвольный набор отношений
  - полностью аксиоматизированную теорию.

# Таксономические отношения



- Варианты названий:
- «принадлежит к классу» («a kind-of» («is-a»))
- Класс - подкласс
- Лингвистика: гипоним – гипероним
- Родовидовое отношение

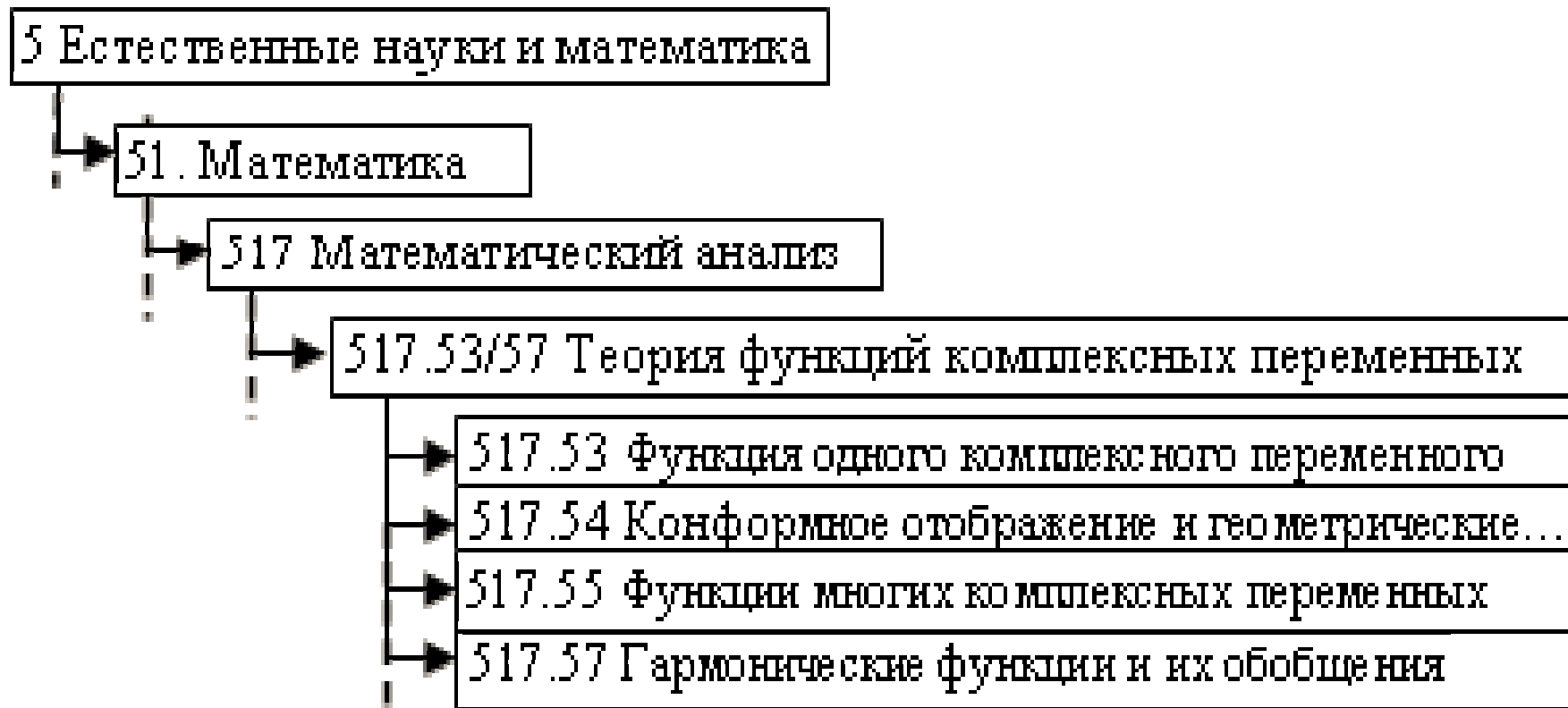
# Свойства таксономических отношений



- Транзитивность:  $A \text{ is\_a } B, B \text{ is\_a } C,$ 
  - $\Rightarrow A \text{ is\_a } C$
- Наследование:
  - $S = \text{свойство } (A)$
  - $B \text{ is\_a } A$
  - $\Rightarrow S = \text{свойство } (B)$



# фрагмент таксономии УДК







# Что такое онтология ?

- На формальном уровне, онтология это система, состоящая из набора понятий и набора утверждений об этих понятиях, на основе которых можно строить классы, объекты, отношения, функции и теории.
- Основные компоненты:
  - Классы или понятия
  - Отношения
  - Функции
  - Аксиомы
  - Примеры



## Типы онтологий:

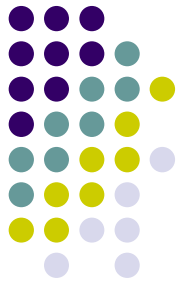
- **верхнего уровня;**
- **предметных областей;**
- **прикладные онтологии (онтологии задач);**
- **лексические онтологии**



# Онтологии верхнего уровня

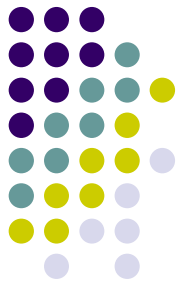
- Cyc's upper ontology  
(<http://www.cyc.com/cyc-2-1/cover.html>)
- Top-level ontologies of universals and particulars (<http://webode.dia.fi.upm.es/>)
- Sowa's top-level ontology  
(<http://www.jfsowa.com/ontology/toplevel.htm>)
- The Standard Upper Ontology(SUMO)  
(<http://suo.ieee.org/>)

# Характеристики онтологий верхнего уровня (ОВУ)



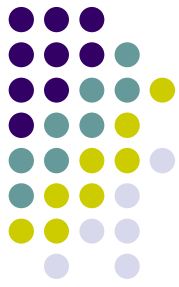
- Набор основных отношений:
  - Иерархические
    - класс-подкласс
    - часть-целое
  - Ассоциативные
- Типичные концепты и принципы разделения:
  - Сущность
  - Явление
  - Объект
  - Процесс
  - Роль
- Нацеленность на многократное повторное использование онтологии

# Онтология СУС



- Lenat D.
- Самый амбициозный проект
- Начат в 1984
- 1 млн. утверждений “common sense”
- Микротеории: пространство, время, причинность
- Онтология 3 тысяч понятий верхнего уровня – в открытом доступе
- [www.cyc.com](http://www.cyc.com)

# Характеристики онтологий предметных областей

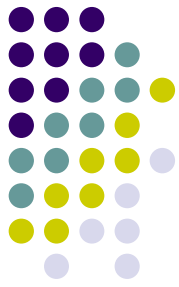


- Существенные отличия от ОВУ:
  - Охват конкретной области знаний (авиация, медицина, культура),
  - Круг решаемых задач и вопросов, на которые онтология отвечает, ограничен выбранной областью,
  - Большой по сравнению с ОВУ объем онтологии,
  - Наличие отношений специфичных для конкретной области.
- Повторное использование только в рамках предметной области

# Онтологии предметной области (библиографические):



- Выражение основных концепций Функциональных требований к библиографическим записям (FRBR) в RDF (<http://purl.org/vocab/frbr/core#>)
- MarcOnt онтология представляющая структуру формата MARC21 и Dublin Core на языке OWL (<http://www.marcont.org/>)
- SKOS семейство формальных языков для описания классификационных схем, тезаурусов, авторитетных файлов. SKOS построенная на RDF и RDFS В настоящее время разработки ведутся в рамках W3C

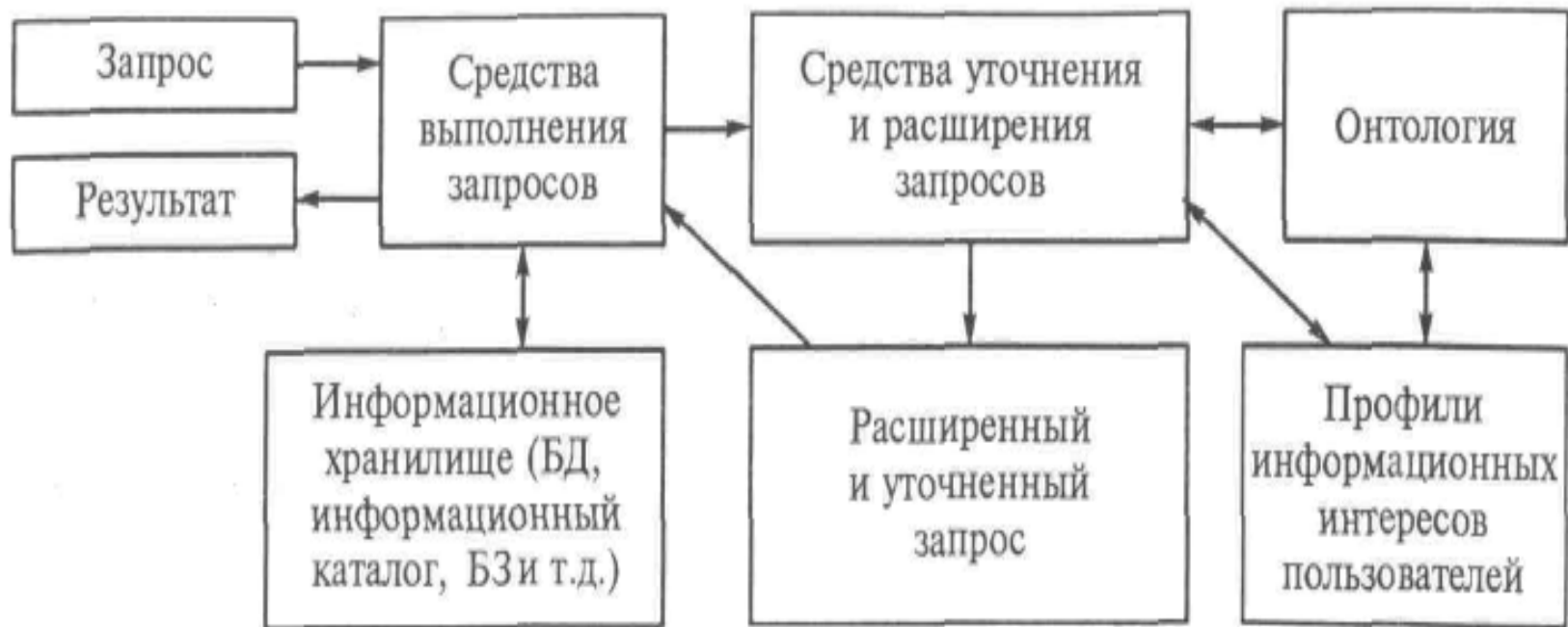
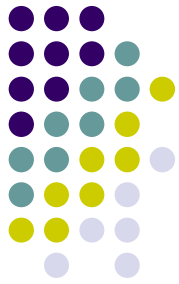


# Применение онтологий

- **Организация эффективного поиска в базах данных, информационных каталогах, базах знаний**



# Организация эффективного поиска в базах данных, информационных каталогах, базах знаний





# Применение онтологий

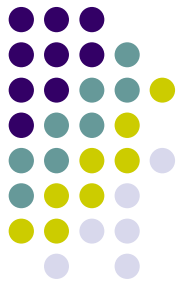
- Семантический поиск в Internet;
- Представление смысла в метаданных об электронных информационных ресурсах;
- Построение и использование баз общих знаний (онтологий верхнего уровня) для различных интеллектуальных систем

# Методики построения онтологии и требования к средствам их спецификации



- *Стандарт онтологического исследования IDEF5* подготовлен фирмой Knowledge Base Systems, Inc. в качестве проекта национального стандарта США. Процесс построения онтологии в рамках IDEF5 состоит из пяти основных этапов
  1. Изучение и систематизация начальных условий. Этот этап устанавливает основные цели и контекст разработки онтологии, а также распределяет роли членов проекта.
  2. Сбор и накопление данных для построения онтологии.
  3. Анализ и группировка собранных данных для облегчения согласования терминологии.
  4. Начальное развитие онтологии. На этом этапе формируется предварительная онтология на основе систематизированных данных.
  5. Уточнение и утверждение онтологии (заключительный этап).

# Онтологический подход к описанию ПрО



Процесс онтологического (концептуального) моделирования предполагает:

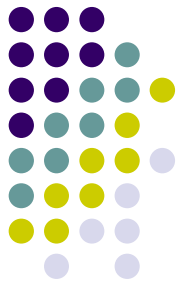
- отбор значимых для рассматриваемой предметной области понятий экспертом предметной области и/или отправляясь от представительного корпуса текстов;
- категоризация (начальная классификация) отобранных понятий;
- дальнейшая систематизация внутри категорий - таксономическая (иерархическая) организация понятийного словаря;
- установление отношений между понятиями в рамках некоторого априори принимаемого ограниченного списка значимых отношений (обычно используются не зависящие от выбора предметной области универсальные отношения: часть - целое, принадлежать к совокупности, быть связанным с, иметь местом и т.п.; могут использоваться и более конкретные, специфичные для данной предметной области отношения).

Иначе говоря, в процессе реконструирования ПрО в виде онтологии необходимо ответить на следующие вопросы:

- Какие объекты существуют в ПрО;
- Какие атрибуты (свойства) имеют объекты и каковы их возможные значения;
- Состоят ли объекты из частей;
- Меняются ли атрибуты и отношения со временем;
- Какие события происходят в различные моменты времени;
- В каких процессах участвуют объекты;
- Каковы причинно-следственные связи событий и состояний.

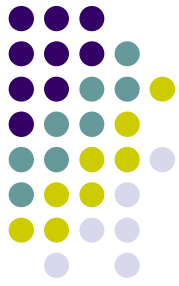
По сути, ответы на эти вопросы определяют возможные виды отношений между понятиями онтологии ПрО.

# Фундаментальные правила разработки онтологии:



- 1) Не существует единственного правильного способа моделирования предметной области – всегда существуют жизнеспособные альтернативы.
- 2) Разработка онтологии – это обязательно итеративный процесс.
- 3) Понятия в онтологии должны быть близки к объектам (физическим или логическим) и отношениям в интересующей вас предметной области. Скорее всего, это существительные (объекты) или глаголы (отношения) в предложениях, которые описывают предметную область.

# Проекты в сфере семантических цифровых библиотек



- **SIMILE** развивает DSpace за счет поддержки произвольных схем и метаданных;
- **Corrib.org** разрабатывает семантическую цифровую библиотеку компонентов, таких, как JeromeDL, MarcOnt, FOAFRealm и HyperCuP;
- **BRICKS** крупнейший проект в области культурного наследия в рамках 6-й программы ЕС применяет технологии семантического веб для управления метаданными в общеевропейской сети цифровых библиотек.
- **FEDORA** гибкая, ориентированная на сферу услуг, архитектура управления и доставки цифрового контента