

Андон Ф.И., Резниченко В.А.

Онтологии – проблемы создания и применения

Институт программных систем НАН Украины

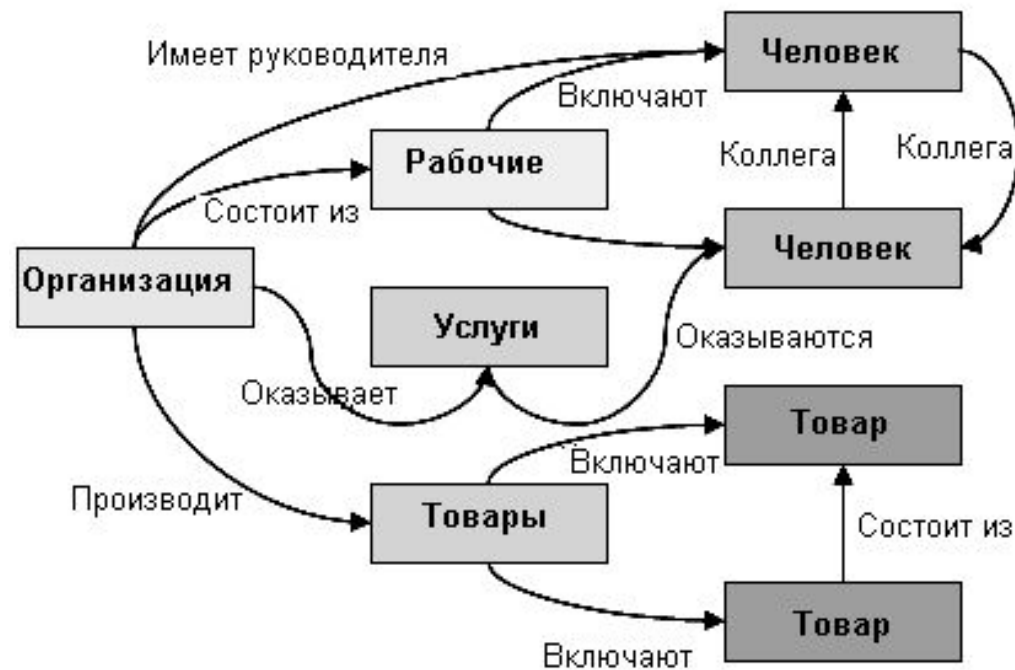
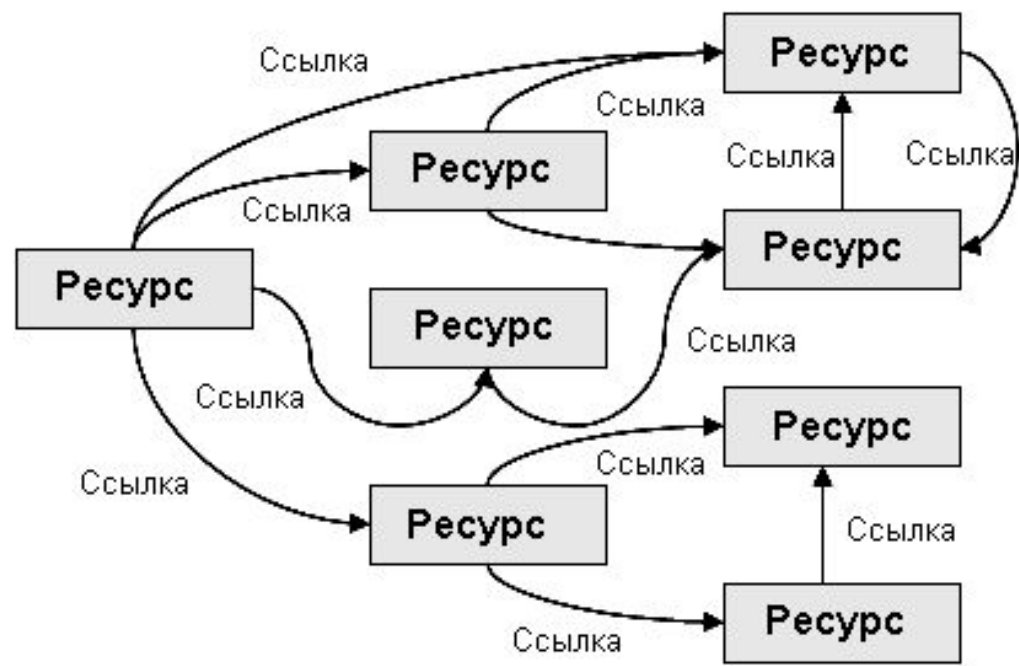
Содержание

1. Базовые понятия
2. Способы придания смысла понятиям
 - управляемые словари
 - таксономии
 - тезаурусы
 - онтологии
3. Онтологии
 - концептуализация
 - спецификация
 - формализация
4. Инженерия онтологий
 - жизненный цикл
 - операции над онтологиями
 - выводы в онтологии
 - Языки онтологий

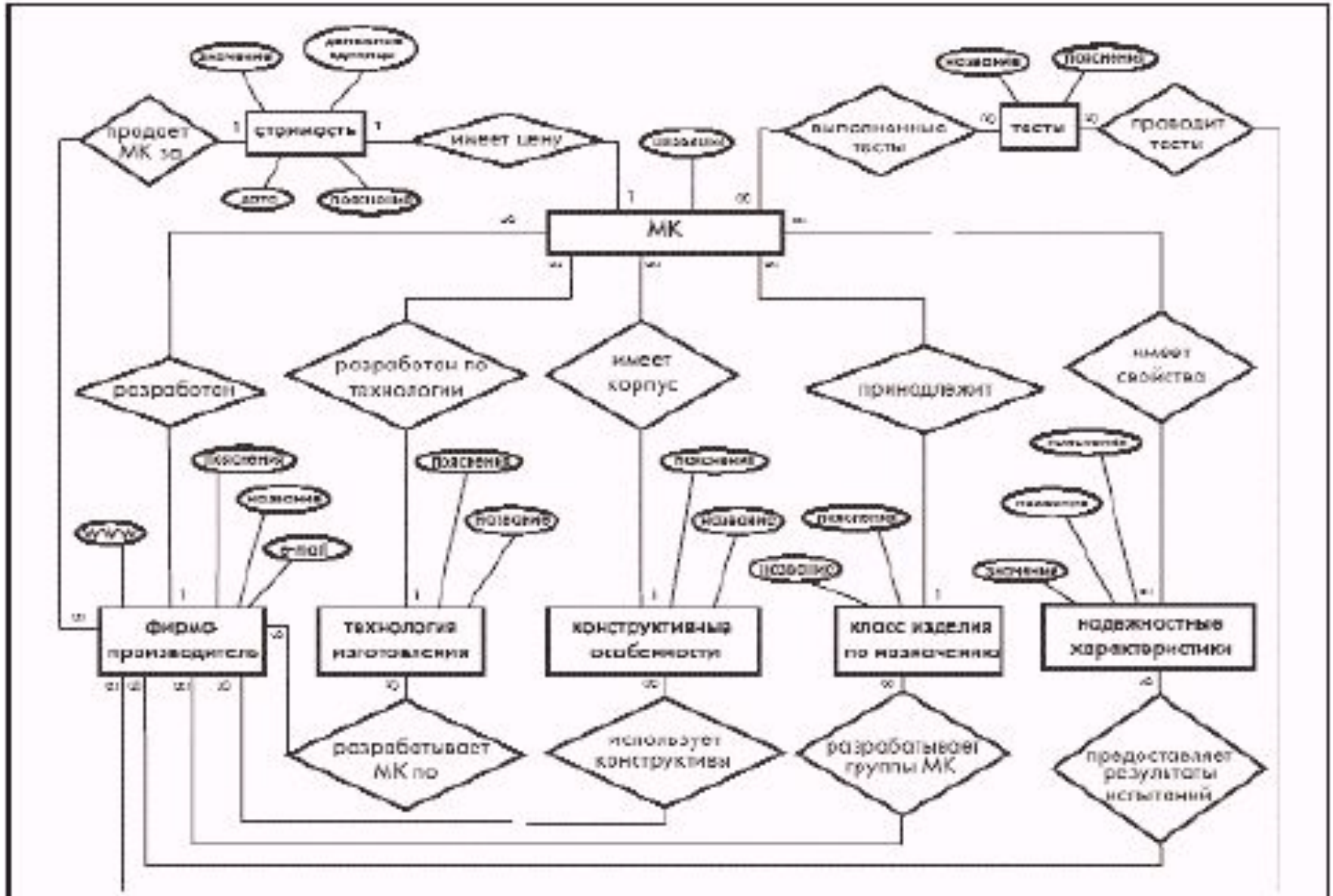
Базовые понятия

- ❖ **Суждение** – форма мысли, в которой утверждается или отрицается что-либо относительно предметов и явлений окружающего мира, их свойств, связей и отношений и которая обладает свойством выражать либо истину или ложь.
- ❖ **Понятие (концепт)** – целостная совокупность суждений, то есть мыслей, в которых что-либо утверждается об отличительных признаках исследуемого объекта, ядром которого являются суждения о наиболее общих и в то же время существенных признаках этого объекта. **Объем (экстенционал)** понятия – класс обобщенных в понятии объектов. **Содержание (интенционал)** понятия – совокупность (обычно существенных) признаков, по которым произведено обобщение и выделение объектов в данном понятии.
- ❖ **Термин** – слово или словосочетание, точно и однозначно именующее понятие и его соотношение с другими понятиями.

Веб и семантический веб



Концептуальная модель предметной области



Инфологическая модель



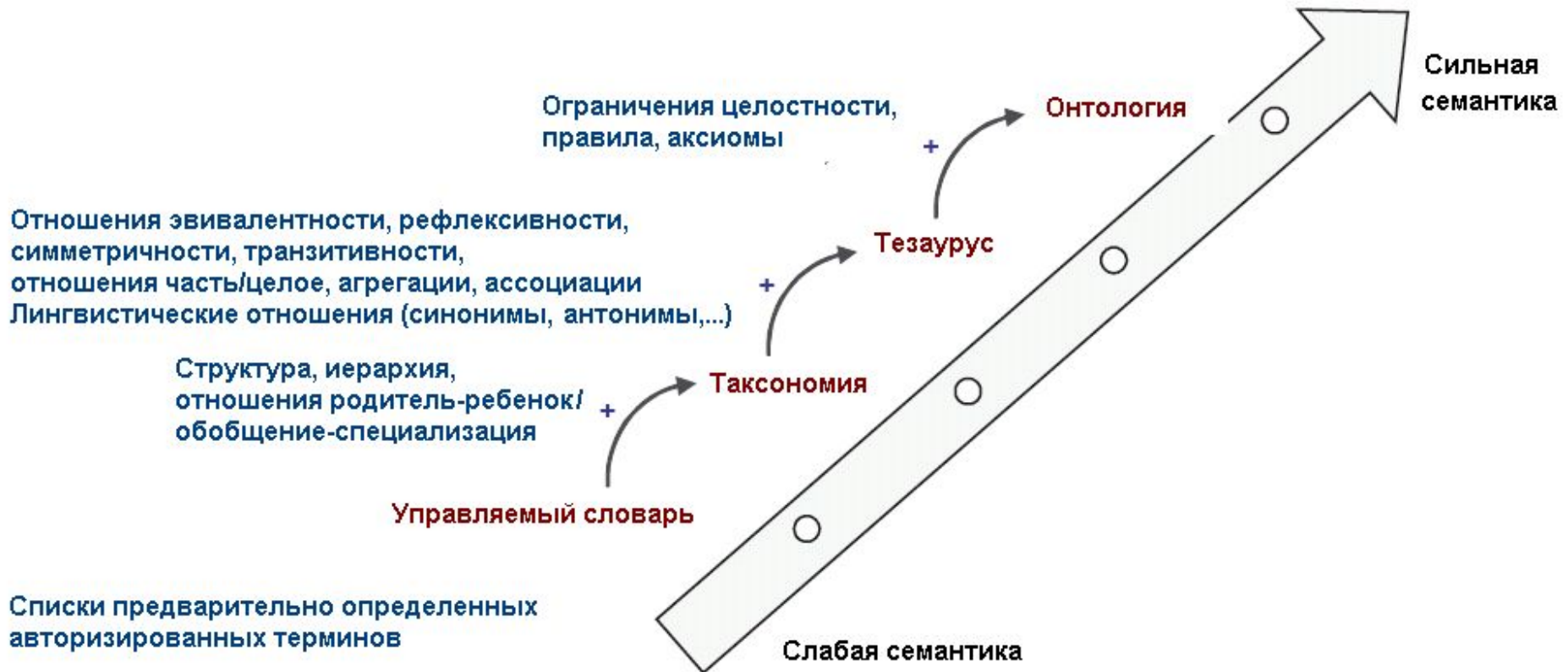
Тезаурус



Онтология



От словаря к онтологии

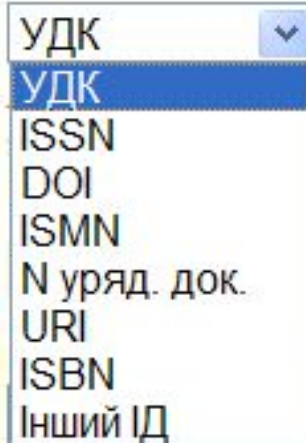


Управляемый словарь (Controlled Vocabulary)

- ❖ **Управляемый словарь** – это способ организации знаний с целью облегчения их представления и последующего поиска
- ❖ Управляемый словарь – это список явно заданных тщательно отобранных терминов (слов, фраз или нотаций). Все термины должны иметь однозначное и неизбыточное толкование (определение).
- ❖ Управляемые словари – это основа классификации.
- ❖ В библиотечных ИС управляемый словарь представляет собой список тщательно отобранных слов и фраз, которые используются для разметки информационных единиц (слов, выражений, документов).
- ❖ Управляемые словари фиксируют возможные варианты выбора значений
- ❖ Управляемые словари уменьшают неопределенность и неоднозначность, присущую естественному языку

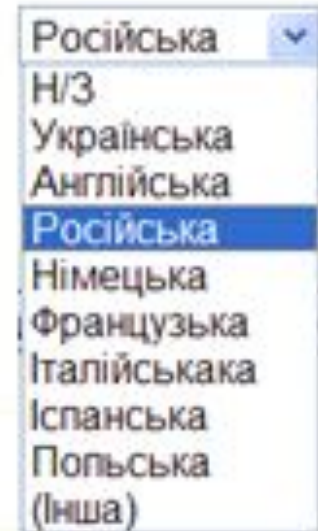
Примеры управляемых словарей

Идентификаторы:



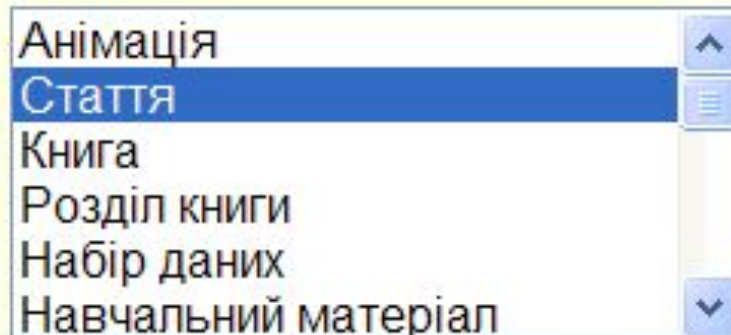
УДК
УДК
ISSN
DOI
ISMN
N уряд. док.
URI
ISBN
Інший ІД

Мова:



Російська
Н/З
Українська
Англійська
Російська
Німецька
Французька
Італійська
Іспанська
Польська
(Інша)

Тип ресурсу:



Анімація
Стаття
Книга
Розділ книги
Набір даних
Навчальний матеріал

Пример управляемого словаря тематических разделов журнала «Доповіді Національної академії наук України»

- ❖ Біологія
- ❖ Біофізика
- ❖ Біохімія
- ❖ Екологія
- ❖ Енергетика
- ❖ Інформатика та кібернетика
- ❖ Інформація
- ❖ Математика
- ❖ Матеріалознавство
- ❖ Медицина
- ❖ Механіка
- ❖ Науки про Землю
- ❖ Теплофізика
- ❖ Фізика
- ❖ Хімія

Примеры использования словарей

- ❖ Библиотечное дело - предметная индексация, предметные рубрики, каталоги, авторитетные файлы
- ❖ Лингвистика - разметка информационных единиц
- ❖ Информационные системы – индексация
- ❖ Программные системы – меню, ниспадающие и всплывающие списки
- ❖ Тематические рубрики в журналах

Таксономия

- ❖ Таксоно́мия (от др.-греч. τάξις — строй, порядок и νόμος — закон) — учение о принципах и практике **классификации и систематизации**.
- ❖ Термин «таксономия» впервые был предложен в 1813 г. для классификации растений и животных и изначально применялся только в биологии.
- ❖ Позже этот термин стал использоваться для обозначения общей теории классификации и систематизации сложных систем как в биологии, так и в других областях знаний: в лингвистике, географии, геологии и т.д.
- ❖ Таксономия – это предметная классификация, которая группирует термины в виде управляемых словарей и упорядочивает их (словари) в виде иерархических структур.

Биологическая иерархическая таксономия, ведущая к человеку:

Фауна: мир животных

Тип: хордовые

Подтип: позвоночные

Класс: млекопитающие

Подкласс: звери

Отряд: приматы

Подотряд: антропоиды

Суперсемейство: гоминоиды

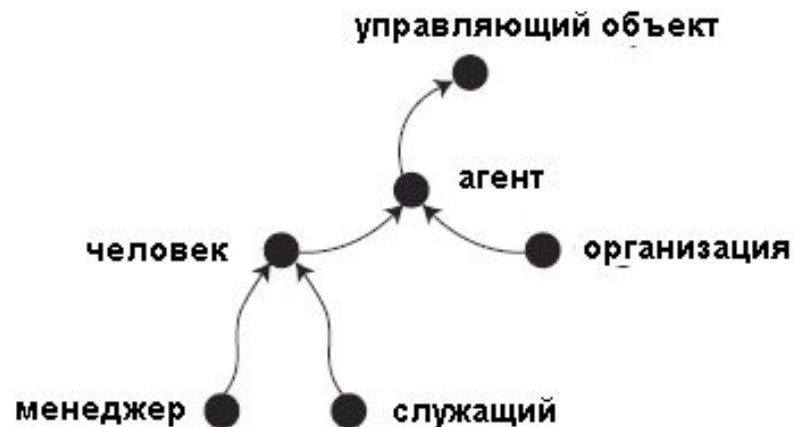
Семейство: гоминиды

Род: Homo

Вид: Sapiens

Класс: диапсиды (рептилии, динозавры, птицы)

Простая иерархическая таксономия «управляющего объекта»:



→ является подклассом

Таксономическая структура

- ❖ Математически таксономией является **древовидная структура** классификаций определенного набора объектов.
- ❖ Вверху этой структуры — объединяющая единая классификация — **корневой таксон** — которая относится ко всем объектам данной таксономии.
- ❖ Таксоны, находящиеся ниже корневого, являются более специфическими классификациями, которые относятся к подмножествам общего множества классифицируемых объектов.
- ❖ Современная биологическая классификация, к примеру, представляет собой иерархическую систему, основание которой составляют отдельные организмы (индивиды), а вершину — один всеобъемлющий таксон; на различных уровнях иерархии между основанием и вершиной находятся таксоны, каждый из которых подчинён одному и только одному таксону более высокого ранга.

Фрагмент таксономії УДК

0 Загальний відділ

00 Загальні питання науки та культури

001 Наука та знання в цілому. Організація розумової праці

002 Документація. Книги. Письменництво. Авторство

003 Системи письма та писемності

004 Комп'ютерна наука та технологія. Застосування комп'ютера

004.2 Комп'ютерна архітектура

004.3 Апаратне забезпечення комп'ютерів

004.4 Програмне забезпечення

004.5 Взаємодія людини і комп'ютера. Інтерфейс користувача

004.6 Дані

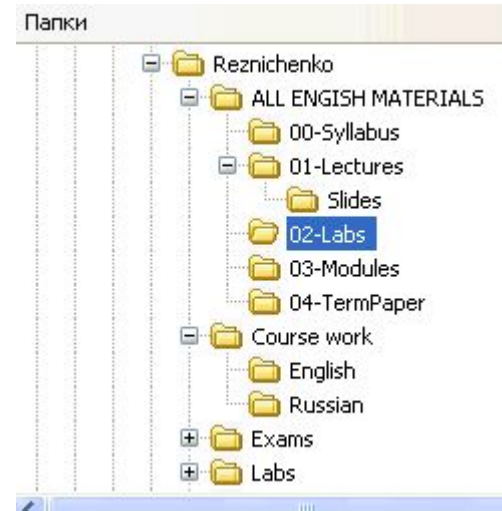
004.7 Комп'ютерні мережі

004.8 Штучний інтелект

004.9 Прикладна техніка, що базується на комп'ютерних системах.

Прикладні інформаційні системи

Иерархическая структура данных и таксономия

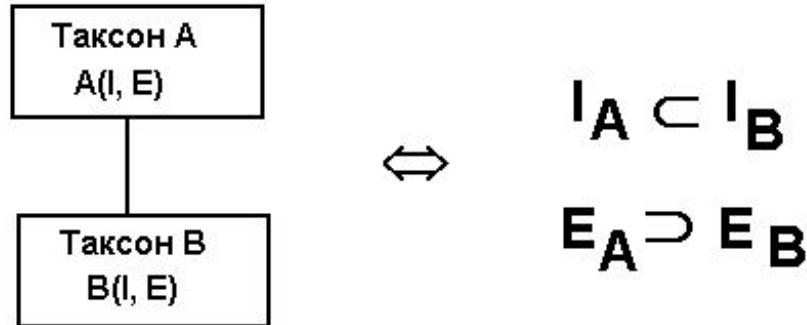


- ❖ В иерархической структуре данных связь понятий означает единственное - наличие между ними взаимосвязи «один ко многим»
- ❖ При этом каждое из понятий, как правило, является самостоятельным в том смысле, например, что «сотрудник» НЕ является ни «организацией» и ни «проектом».
- ❖ В каждой организации работает множество сотрудников и имеется множество проектов, но каждый сотрудник работает в одной организации и каждый проект выполняется в одной организации.
- ❖ Другим примером такой иерархии является структура папок в операционной системе компьютера.

Суть таксономической иерархического отношения

E – экстенционал – множество индивидов, принадлежащих таксону

I – интенционал – множество свойств, характеризующих индивиды экстенционала таксона



Таксономические отношения также называются:

- ❖ отношением «обобщения/специализации»,
- ❖ «родовидовым» отношением
- ❖ отношением «супертип/подтип»
- ❖ отношением «суперкласс/подкласс»
- ❖ В англоязычной литературе также говорится об отношении «Is a»

Тезаурус

- ❖ **Тезаурус** (от греч. thesaurós — сокровище, сокровищница),
- ❖ В широком смысле **Тезаурус** интерпретируются как описание системы знаний о действительности.
- ❖ Тезаурус является расширением таксономии в том смысле, что в тезаурусе помимо родо-видовых отношений могут существовать любые другие отношения, которые на множестве понятий формируют сложную сетевую структуру.
- ❖ Тезаурусы, особенно в электронном формате, являются одним из действенных инструментов для описания отдельных предметных областей.

Тезаурус в лингвистике

- ❖ С точки зрения лингвистики тезаурус - это множество смысловых единиц некоторого языка с заданной на нём системой семантических отношений. Тезаурус фактически определяет семантику языка (национального языка, языка конкретной науки или формализованного языка для автоматизированной системы управления).
- ❖ Например, лингвистический, тезаурус содержит:
 - ❖ Морфологические и синтаксические свойства (часть речи, род, склонение, корень, словоформы в различных падежах, родах и числах)
 - ❖ Семантика (значение, синонимы, антонимы, гиперонимы, гипонимы)
 - ❖ Родственные слова
 - ❖ Происхождение
 - ❖ Фразеологизмы и устойчивые сочетания
 - ❖ Перевод на другие языки

Пример тезауруса слова «замок» в Викисловаре

за́мок

Морфологические и синтаксические свойства

за́-мок

Существительное, неодушевлённое, мужской род, 2-е склонение (тип склонения 3*а по классификации А. Зализняка).

Приставка: **за-**; корень: **-мок-**.

Произношение

МФА: ед. ч. [ˈza.mək], мн. ч. [ˈzam.kɨ]  [Пример произношения](#)

Семантические свойства

Значение

- здание (или комплекс зданий), обычно обнесённое стеной и сочетающее в себе оборонительную и жилую функции ◆ Почтенный **за́мок** был построен, // Как **замки** строиться должны: // Отменно прочен и спокоен // Во вкусе умной старины. *А. С. Пушкин, «Евгений Онегин»*

Синонимы

- частичн.: дворец, крепость

Антонимы

- лачуга

Гиперонимы

- сооружение, здание, крепость, жильё

Гипонимы

Родственные слова

Ближайшее родство:

[\[показать\]](#)

Этимология

Происходит от польск. *zamek*, которое через чешск. *zátek* калькирует ср.-в.-нем. *slōz* «замок, запор», затем «перемычка долины, оврага и под». Последнее в свою очередь является калькой лат. *clōsa* «запор, форт, укрепление». Известно с XVII в. (Беседа трёх святит). Связано с замыкать (замкнуть). *Использованы данные словаря М. Фасмера; см. Список литературы.*

Фразеологизмы и устойчивые сочетания

- феодальный замок
- воздушный замок; строить воздушные замки

Перевод

[\[показать\]](#)

падеж	ед. ч.	мн. ч.
Им.	за́мок	за́мки
Р.	за́мка	за́мков
Д.	за́мку	за́мкам
В.	за́мок	за́мки
Тв.	за́мком	за́мками
Пр.	за́мке	за́мках



Замок



Информационно-поисковые тезаурусы

- ❖ В 70-х гг. 20 в. получили распространение информационно-поисковые тезаурусы. В этих тезаурусах выделены специальные лексические единицы — дескрипторы, по которым можно осуществлять автоматический поиск документальной информации. С каждым словом такого тезауруса сопоставляется дескриптор (дескрипторная статья),

Дескрипторная статья может иметь следующую структуру:

- ❖ заглавный дескриптор;
- ❖ ключевые слова из класса эквивалентности;
- ❖ дескрипторы, подчиняющие заглавный;
- ❖ дескрипторы, подчиненные заглавному;
- ❖ дескрипторы, ассоциированные с заглавным.
- ❖ Для дескрипторов явным образом указываются семантические отношения: род - вид, часть - целое, ассоциации (причина - следствие; сырье - продукт; процесс - объект; процесс — субъект; свойство — носитель свойства).

ANSI/NISO Monolingual Thesaurus Standard

ANSI/NISO Monolingual Thesaurus Standard (NISO - National Information Standards Organization) определяет тезаурус как

- ❖ «упорядоченный управляемый словарь, структурированный таким образом, что в нем между терминами четко определены и идентифицированы отношения эквивалентности, гомографии, иерархии и ассоциации с использованием стандартизированных индикаторов этих отношений... Первичная задача тезаурусов заключается в том, чтобы облегчить поиск документов и достигнуть согласованности в выполнении индексации письменных или другим способом полученных документов».
- ❖ Таким образом, согласно этого определения» в тезаурусах определяется четыре различных типа связей: **эквивалентность, омонимия, иерархия, ассоциация**

Отношения и их эквиваленты	Смысл	Примеры
эквивалентность, синоним, аналогично, используется вместо	Термин X имеет тот же или почти тот же смысл, что и термин Y	«электронная библиотека» является синонимом для «цифровая библиотека»
омоним, имеет такое же написание, гомография	Термин X имеет такое же написание, что и Y, но они имеют различный смысл	Закреть замок на замок, чтобы замок не замок
шире, чем иерархия: является родительским для	Термин X шире по смыслу, чем Y	“Организация” шире по смыслу, чем «финансовое учреждение»
уже, чем иерархия: является дочерним для	Термин X уже по смыслу, чем Y	«финансовое учреждение» ”уже по смыслу, чем «организация»
ассоциация, ассоциируется с, связан с, используется для	Термин X ассоциируется с термином Y. Существует неспецифицируемая связь между ими.	«гвоздь» связан с «молотком»

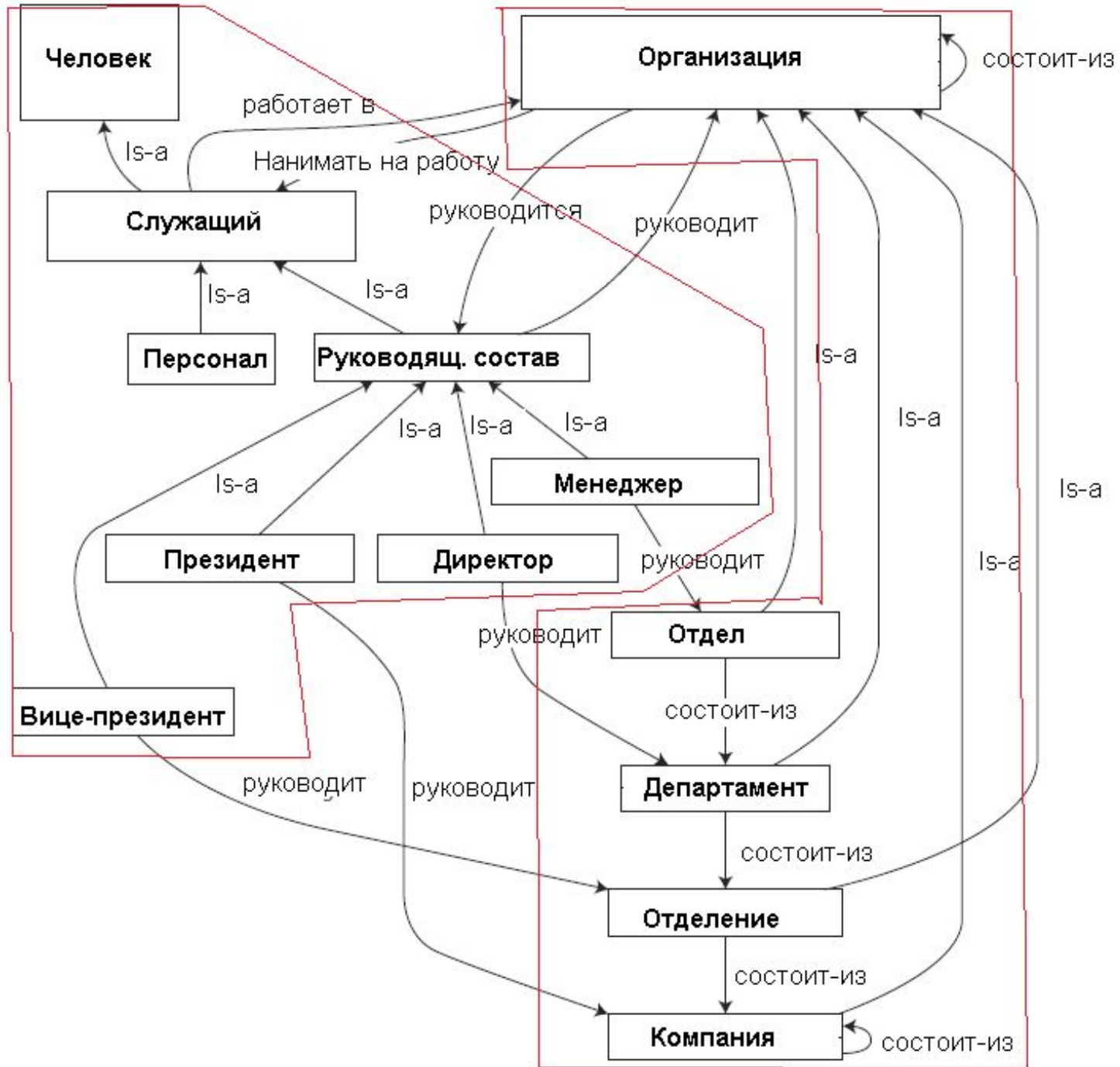
Пример: тезаурус как связь двух таксономий



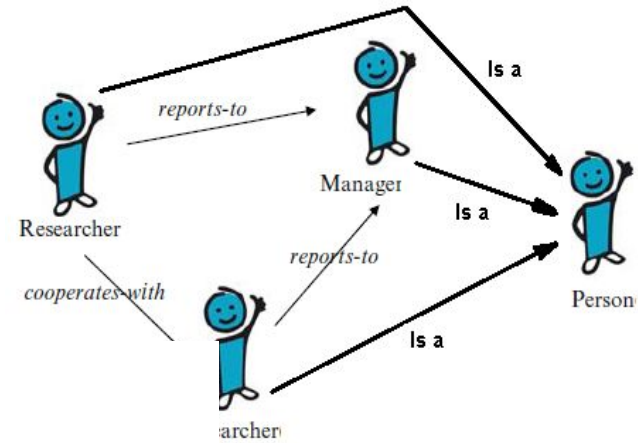
Пример фрагмента тезауруса для понятия «успеваемость»

Связь	Термин (понятие)
Используется для	Средний балл Оценка знаний Оценка преподавания
Более узкое понятие	Успеваемость в школе Успеваемость в колледже Успеваемость в ВУЗе Успеваемость по математике
Более широкое понятие	Успех
Связан с	Мотивация успеваемости Прогнозирование успеваемости Способности к обучению Образование Учебный процесс

Таксономии, тезаурусы и онтологии



Тезаурусы и онтологии



Правила переходов для свойства «Семейное положение»



Два **ИССЛЕДОВАТЕЛЯ** **сотрудничают**, если

- 1) они имеют общую цель и
- 2) предпринимают определенные действия для достижения этой цели

Назначение словарей, таксономий, тезаурусов, онтологий

- ❖ Это основа для выработки единой, согласованной, нормативной, однозначно понимаемой, полной и непротиворечивой терминологии, используемой всеми, кто имеет отношение к ПО.
- ❖ Это средства, предназначенные для классификации, структурирования, систематизации, моделирования и придания смысла понятиям и связям, относящимся к ПО.
- ❖ Это средства для “исчерпывающего” описания информационной модели ПО, включая и ее семантику, самостоятельно, то есть не зависимо от задач (оперативных знаний), которые будут решаться с использованием этой модели.
- ❖ Это основа для коммуникации между людьми, компьютерными агентами и людьми и компьютерными агентами.

Определения онтологий

N. Guarino and P. Giaretta. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In N. Mars, editor, *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, pages 25–32. IOS Press, Amsterdam, 1995.

Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем. Учебник. — СПб.: Питер, 2000. 382 с.

- 1) **Онтология как философская дисциплина.** (ὄντος — сущее, то, что существует и λόγος — учение, наука)
- 2) Онтология как неформальная концептуальная система.
- 3) Онтология как формальный взгляд на семантику.
- 4) **Онтология как формальная, явная спецификация согласованной концептуализации.**
- 5) Онтология как представление концептуальной системы через логическую теорию, характеризуемую: а) специальными формальными свойствами или б) только ее назначением.
- 6) Онтология как словарь, используемый логической теорией.
- 7) Онтология как (метауровневая) спецификация логической теории.

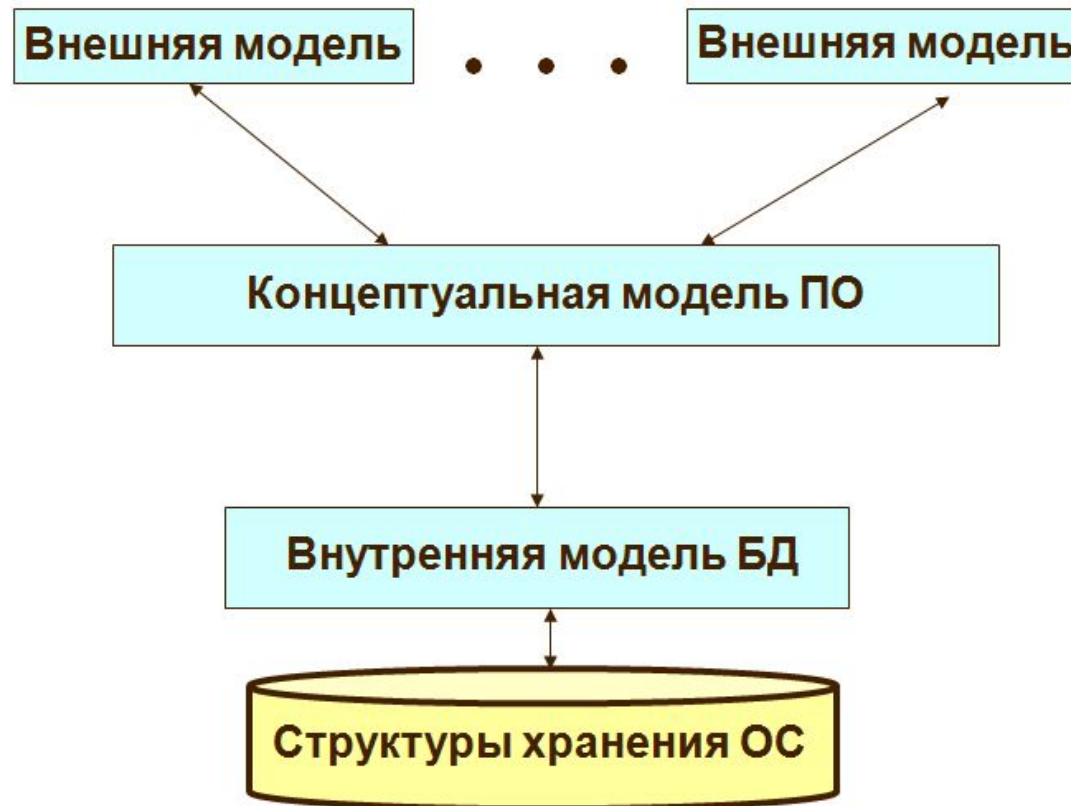
Компоненты онтологий

- ❖ **Индивиды** (объекты, элементы, экземпляры, термины)
- ❖ **Классы** (множества, коллекции, типы, унарные отношения, понятия)
- ❖ **Атрибуты** - аспекты, свойства, характеристики, особенности, параметры, которые могут быть присущими индивидам
- ❖ **Отношения (связи)** – способ выражения взаимосвязи между индивидами
- ❖ **Аксиомы** (ограничения, правила) – выражение суждений, которые имеют место в рассматриваемой предметной области
- ❖ **Правила вывода** (умозаключение) – логическое действие, в результате которого из одного или нескольких известных и определенным образом связанных суждений получается новое суждение (дедукция, индукция, традиция, вывод по аналогии)
- ❖ **Функции** (процессы, процедуры,...)

Что такое концептуализация? – концептуализация в базах данных

ANSI/X3/SPARC Study Group on Data Base Management Systems: (1975), *Interim Report. FDT*, ACM SIGMOD bulletin. Volume 7, No. 2

ISO TC97/SC5/WG3. Concepts and terminology for the conceptual schema and the information base (edited by J.J. van Griethuysen) March 15, 1982, 150 p.



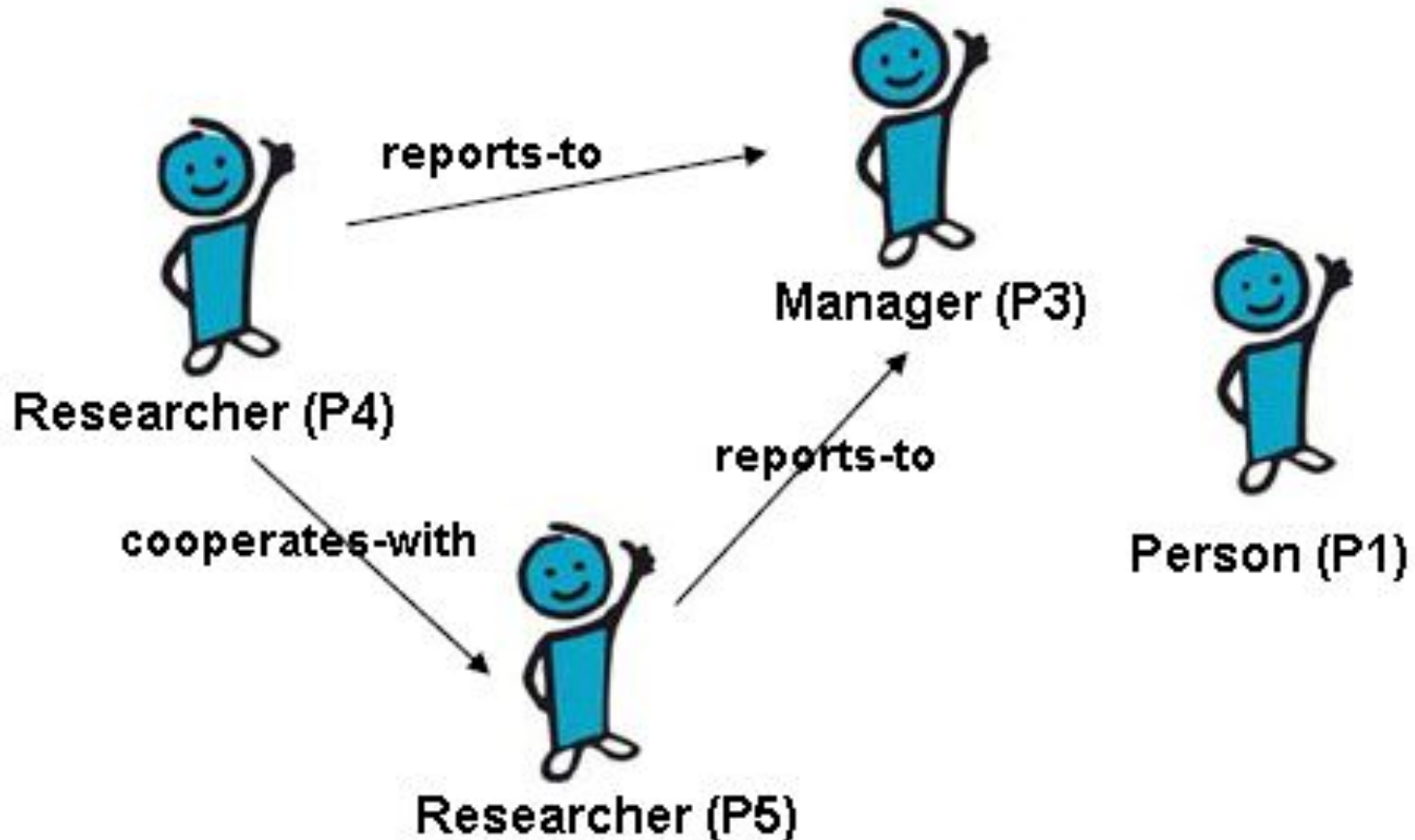
Что такое концептуализация? – концептуализация в ИИ

M.R. Genesereth and N. J. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann, Los Altos, CA, 1987.

«Основой формального представления знаний является **концептуализация**, представляющая собой множество объектов, понятий и других сущностей, которые, как предполагается, существуют в некоторой представляющей интерес предметной области, и имеющиеся между ними отношения. Концептуализация является абстрактным упрощенным взглядом на мир, который мы хотим представить для каких-то целей. Каждая база знаний, система, базирующаяся на знаниях, агент знаний прямо или косвенно используют концептуализацию».

Используемый пример

- ❖ наша предметная область состоит из множества лиц (Person);
- ❖ среди них выделяются руководители (Manager) и исследователи (Researcher);
- ❖ исследователи отчитываются (reports-to) перед руководителями;
- ❖ исследователи могут сотрудничать (cooperates-with) между собой



Формальное определение концептуализации - экстенсиональная реляционная структура

Определение 1. Экстенсиональная реляционная структура – это кортеж (D, R) , где:

- D – множество, называемое доменом (универсумом)
- R – множество отношений на D

Здесь R – обычные математические отношения различной арности, то есть множества кортежей элементов из D

Пример

$D = \{P1, P2, .. \}$

$R = \{\text{Person, Manager, Researcher, cooperates-with, reports-to}\}$

Формальное определение концептуализации – система, мир, состояние мира

- ❖ Под **системой** мы будем подразумевать часть реального мира, которая представляет для нас интерес с точки зрения процесса моделирования, и которая, при заданном уровне детализации, воспринимается наблюдающим ее агентом (который обычно является внешним по отношению к самой системе) в виде совокупности «наблюдаемых переменных».
- ❖ **Определение 2. (мир и состояния мира).**
По отношению к конкретной моделируемой системе S , **состоянием мира** для S является уникальное присвоение значений наблюдаемым переменным, которые характеризуют систему. **Мир** – это линейно упорядоченное множество (*linearly ordered set*) состояний мира, которые отражают эволюцию системы во времени. Если для простоты абстрагироваться от времени, то мир совпадает с состоянием мира.

Формальное определение концептуализации – интенциональное отношение и концептуализация

❖ **Определение 3 (интенциональное отношение или концептуальное отношение).**

Пусть S является произвольной системой, D является произвольным множеством выделенных элементов S , а W – множество состояний мира для S (также называемыми мирами или возможными мирами).

Интенциональным отношением (или **концептуальным отношением**) ρ^n n -арности n на $\langle D, W \rangle$ является всюду определенная функция $\rho^n : W \rightarrow 2^{D^n}$ из множества W в множество всех n -арных экстенциональных отношений на D .

❖ **Определение 4. (интенциональная реляционная структура или концептуализация).**

Интенциональная реляционная структура (или **концептуализация**)

– это тройка (D, W, \mathfrak{R}) где:

- D – домен (универсум)
- W – множество возможных миров
- \mathfrak{R} – множество интенциональных (концептуальных) отношений

Пример концептуализации

- $D = \{P1, P2, \dots\}$ – универсум
- $W = \{w1, w2, \dots\}$ – множество возможных миров
- $\mathfrak{R} = \{\text{PERSON, MANAGER, RESEARCHER, COOPERATES-WITH, REPORTS-TO}\}$ – множество концептуальных отношений

Если предположить, что унарные отношения не изменяются (являются постоянными во всех состояниях мира) и множество людей совпадает с универсумом, а бинарные отношения изменяются, то можно таким образом специфицировать интенциональные отношения (как функции)

- во всех мирах w из W : $\text{PERSON}(w) = D$
- во всех мирах w из W : $\text{MANAGER}(w) = \{\dots, P3, \dots\}$
- во всех мирах w из W : $\text{RESEARCHER}(w) = \{\dots, P4, \dots, P5, \dots\}$
- $\text{REPORTS-TO}(w1) = \{\dots, (P4, P3), (P5, P3), \dots\}$
- $\text{REPORTS-TO}(w2) = \{\dots, (P4, P3), (P5, P3), (P17, P3), \dots\}$
- $\text{REPORTS-TO}(w3) = \dots$
- $\text{COOPERATES-WITH}(w1) = \{\dots, (P4, P5), \dots\}$
- $\text{COOPERATES-WITH}(w2) = \dots$

Спецификация – экстенсиональная и интенциональная

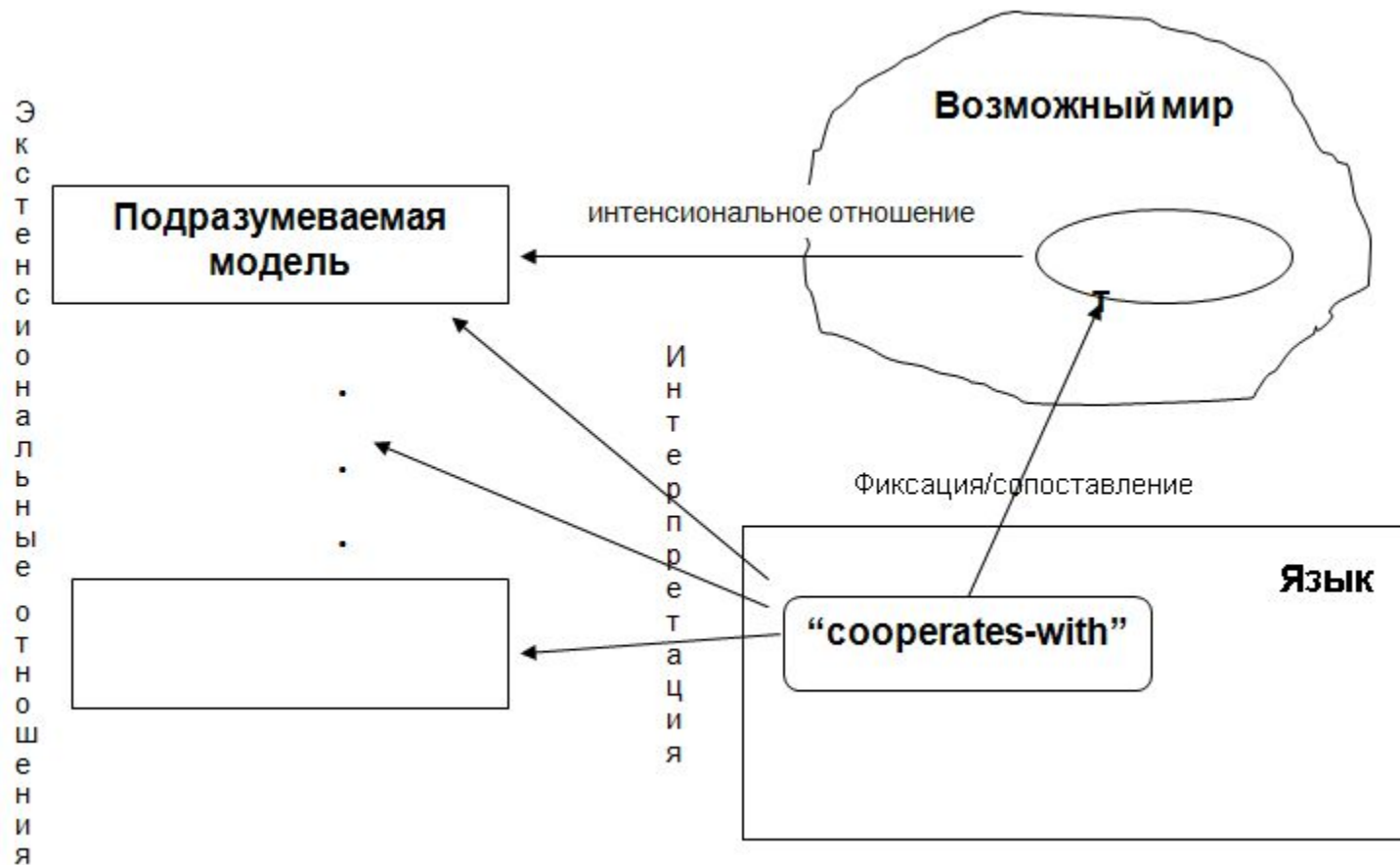
- ❖ **Экстенсиональная спецификация** концептуализации заключается в перечислении всех экстенсионалов (экстенсиональных отношений) каждого из концептуальных отношений для всех состояний возможных миров.
- ❖ **Интенциональная спецификация** – это выбор языка и ограничения его интерпретации интенциональным способом путем определения в этом языке набора подходящих аксиом (правил, ограничений, смысловых постулатов).

Например, мы могли бы сформулировать аксиомы, что:

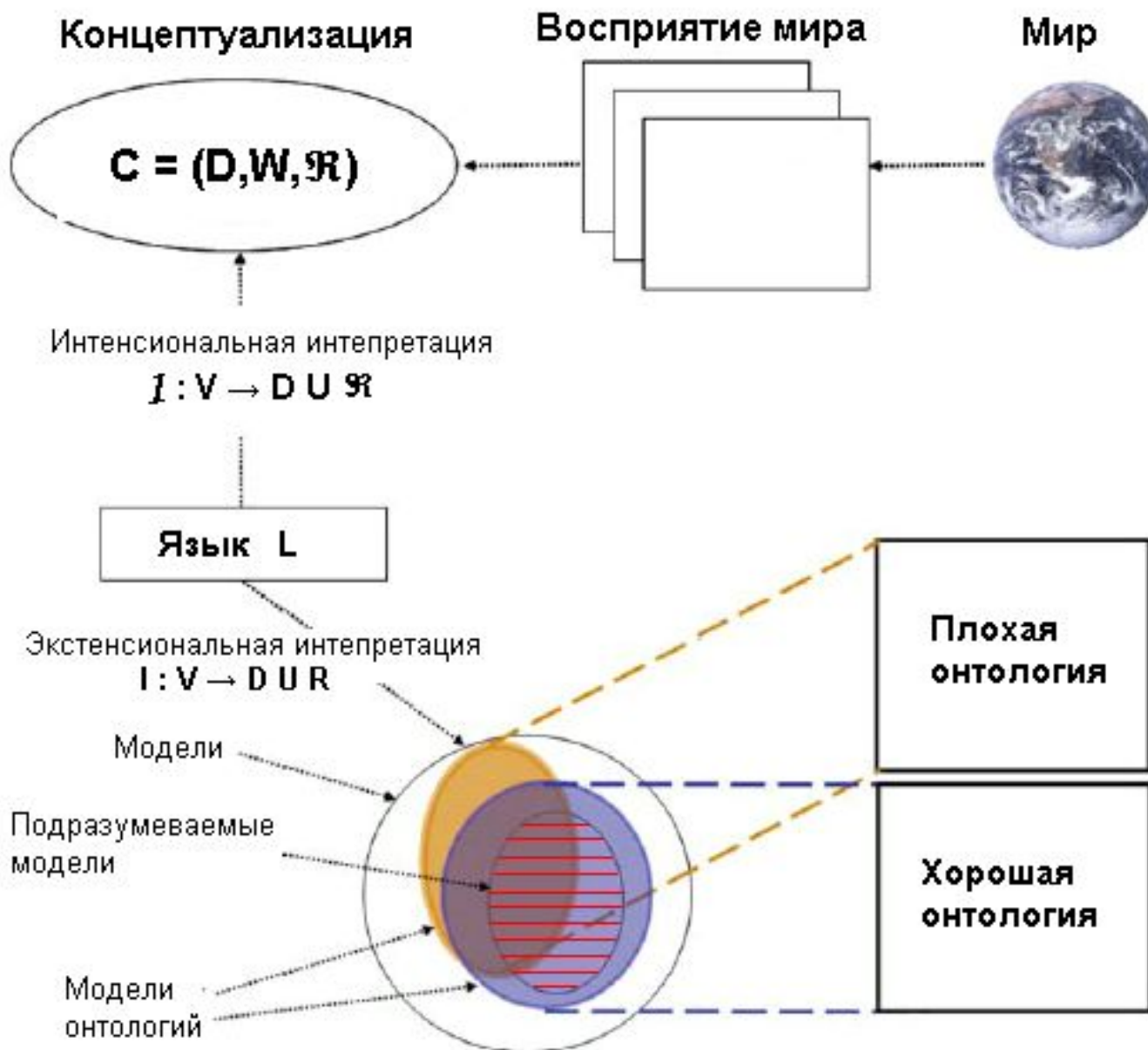
- *reports-to* является ассиметричным и нетранзитивным, а
- *cooperates-with* – симметричным и иррефлексивным.

В этом случае онтология – это просто множество таких аксиом, то есть это логическая теория, предназначенная для представления **только** подразумеваемых моделей, соответствующих определенной концептуализации.

Спецификация – подразумеваемая модель



Мир, концептуализация, язык, интерпретация, онтологии



Спецификация - Онтологическое сопоставление

Определение 5 (экстенсиональная структура первого порядка / модель).

Пусть L – логический язык первого порядка со словарем V и $S = (D, R)$ – экстенсиональная реляционная структура.

Экстенсиональная структура первого порядка (также называемая **моделью** L), это кортеж $M = (S, I)$, где I называется функцией экстенсиональной интерпретации и представляет собой всюду определенную функцию $I : V \rightarrow D \cup R$, которая отображает символы словаря V в D или R .

Определение 6 (интенсиональная структура первого порядка / онтологическое сопоставление)

Пусть L – логический язык первого порядка со словарем V и $C = (D, W, \mathfrak{R})$ – концептуализация.

Интенсиональная структура первого порядка (также называемая **онтологическим сопоставлением**), это кортеж $K = (C, I)$, где I называется функцией интенсиональной интерпретации и представляет собой всюду определенную функцию $I : V \rightarrow D \cup \mathfrak{R}$, которая отображает символы

Подразумеваемая модель и онтология

Определение 7 (Подразумеваемая модель)

Пусть $C = (D, W, \mathfrak{R})$ является концептуализацией, пусть L – логический язык первого порядка со словарем V и с онтологическим сопоставлением $K = (C, I)$. Модель $M = (S, I)$ с экстенциональной реляционной структурой $S = (D, R)$, называется подразумеваемой моделью языка L согласно K тогда и только тогда, когда:

- Для всех символов-констант $c \in V$ имеем $I(c) = I(c)$
- Для любого состояния мира $w \in W$ выполняется следующее: для каждого предикатного символа $v \in V$ существует интенциональное отношение $\rho \in \mathfrak{R}$ такое, что $I(v) = \rho$ и $I(v) = \rho(w)$

Определение 8 (Онтология)

Пусть $C = (D, W, \mathfrak{R})$ является концептуализацией, пусть L – логический язык первого порядка со словарем V и с онтологическим сопоставлением $K = (C, I)$. Онтология O_K – это **логическая теория**, состоящая из множества таких формул L , что получаемые при этом модели насколько это возможно приближаются к подразумеваемым моделям L согласно K .

Пример онтологии

1. Таксономические правила, указывающие, что исследователи и руководители являются видовыми понятиями человека

$$O_1 = \{\text{Researcher}(x) \rightarrow \text{Person}(x), \text{Manager}(x) \rightarrow \text{Person}(x)\}$$

2. Введем правила, определяющие области допустимых значений для бинарных отношений

$$O_2 = O_1 \cup \{\text{cooperates-with}(x, y) \rightarrow \text{Person}(x) \ \& \ \text{Person}(y), \\ \text{reports-to}(x, y) \rightarrow \text{Person}(x) \ \& \ \text{Person}(y)\}$$

3. Отношение *cooperates-with* является симметричным отношением

$$O_3 = O_2 \cup \{\text{cooperates-with}(x, y) \leftrightarrow \text{cooperates-with}(y, x)\}$$

4. Отношение *reports-to* является транзитивным

$$O_4 = O_3 \cup \{\text{reports-to}(x, y) \ \& \ \text{reports-to}(y, z) \rightarrow \text{reports-to}(x, z)\}$$

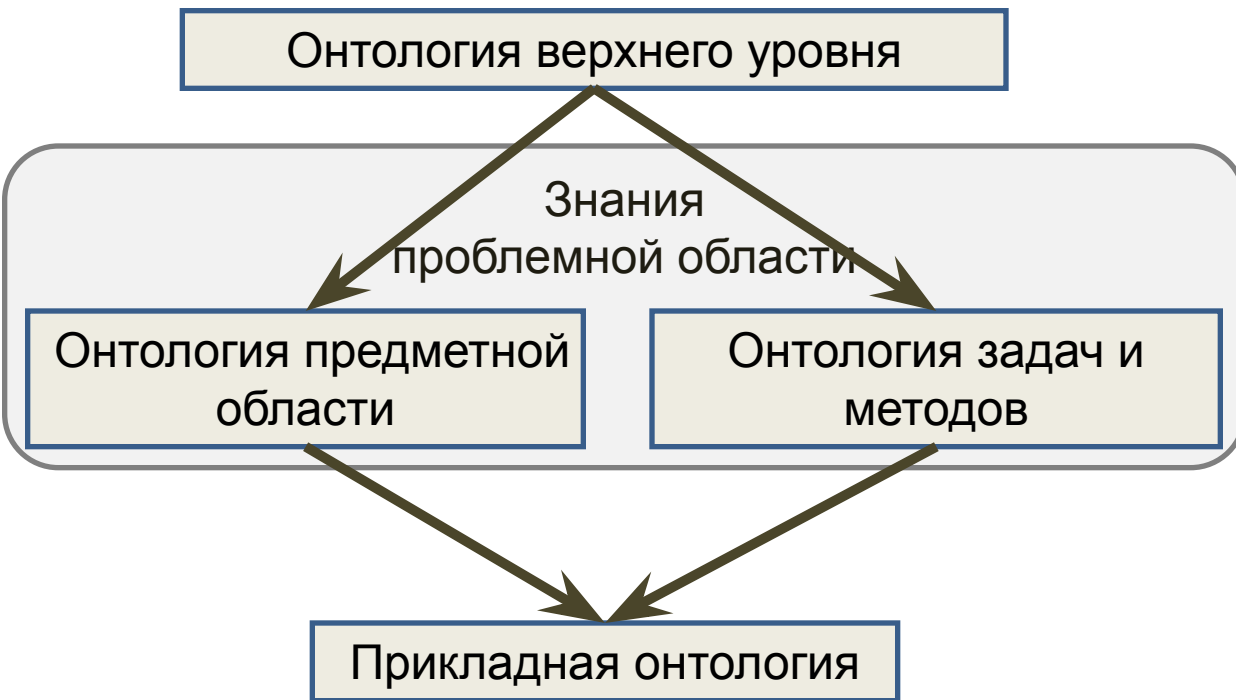
5. Непересекаемость исследователей и руководителей

$$O_5 = O_4 \cup \{\text{Manager}(x) \rightarrow \neg \text{Researcher}(x)\}$$

Инженерия онтологий - определение

- ❖ **Инженерия программного обеспечения** — приложение систематического, дисциплинарного, измеримого подхода к развитию, оперированию и обслуживанию программного обеспечения, а также исследованию этих подходов; то есть, приложение дисциплины инженерии к программному обеспечению.
- ❖ **Инженерия онтологий** в компьютерных науках – это дисциплина, которая изучает методы, методологии и средства построения и использования онтологий.
- ❖ Инженерия онтологий направлена на явное извлечение и представление знаний, содержащихся в предметных областях различного назначения.
- ❖ Инженерия онтологий направлена на решение проблем интероперабельности, вызванных семантическими проблемами, то есть проблемами возможного неоднозначного понимания терминов той или иной предметной области.
- ❖ Инженерия онтологий включает следующие аспекты:
 - жизненный цикл онтологий;
 - процесс разработки онтологий;
 - методы и методологии построения онтологий;
 - набор инструментов и языков для их построения и поддержки

Классификация онтологий по содержанию

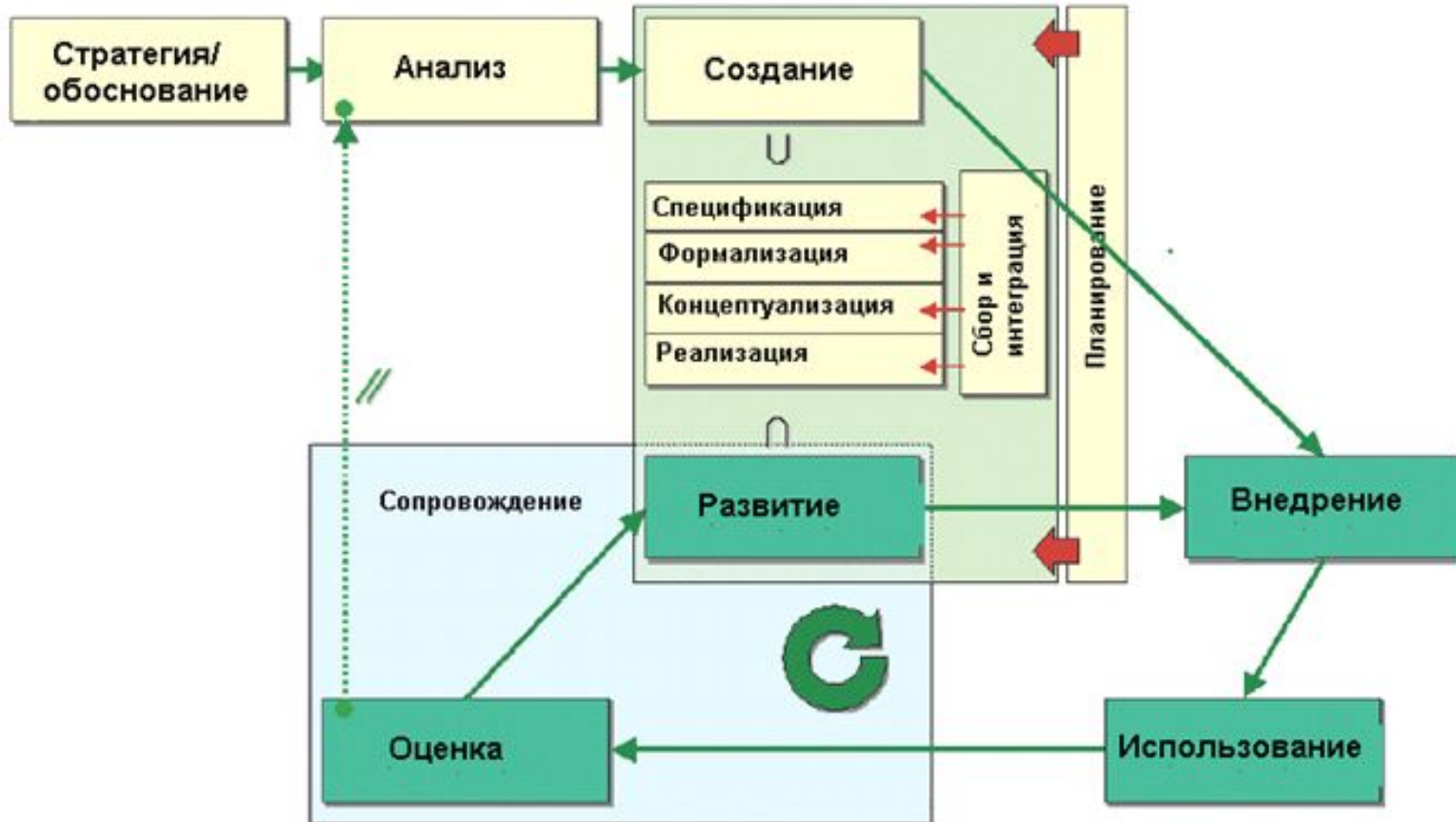


Сущности, связи, время, пространство, события

Составление расписаний, резервирование билетов, управление производством, компьютерные науки

Конкретные онтологии, например, онтология Института программных систем

Инженерия онтологий – жизненный цикл



Стратегия:

- Для каких целей создается онтология?
- Какую область будет охватывать онтология?
- Возможность повторного использования существующих онтологий
- Кто и как будет использовать и поддерживать онтологию?

Анализ – это, в частности: - извлечение знаний (knowledge discovery);
- интеллектуальный анализ данных (data mining)

Инженерия онтологий – методология создания

- ❖ Четкая спецификация ПО
- ❖ Построение **гlossария терминов** (терминов понятий, атрибутов, связей и пр.) с предоставлением их определений на естественном языке и указанием их синонимов, акронимов и других характеристик.
- ❖ Выявление и четкое и достаточно глубокое определение **понятий**. Это основа для создания управляемых словарей.
- ❖ Выявление и четкое определение множества **свойств** (атрибутов, характеристик), характеризующих каждое понятие. Обязательное выделение множества обязательных свойств и свойств, идентифицирующих понятия. Спецификация допустимых **значений** свойств и других ограничений на значения свойств.
- ❖ Выявление и четкое определение **родовидовых зависимостей** (связей) между понятиями и, тем самым построение таксономий на множестве понятий. Определение таких свойств таксономических связей, как непересекаемость декомпозиции и полнота декомпозиции
- ❖ Выявление и четкое определение других **произвольных бинарных связей** между понятиями с указанием к какому типу связей они относятся (эквивалентность, часть/целое, агрегация, ассоциация, причина-следствие и т.д.). Необходимо указать **мощность** окончаний связей (1:1, 1: M, M:N) и их **факультативность** (обязательная, факультативная). Тем самым строится тезаурус.
- ❖ Выявление и четкое определение **аксиом** (правил, прикладных ограничений), характеризующих углубленную семантику понятий, атрибутов и связей. Тем самым строится онтология.
- ❖ Построение экстенсионалов.

Инженерия онтологий – операции

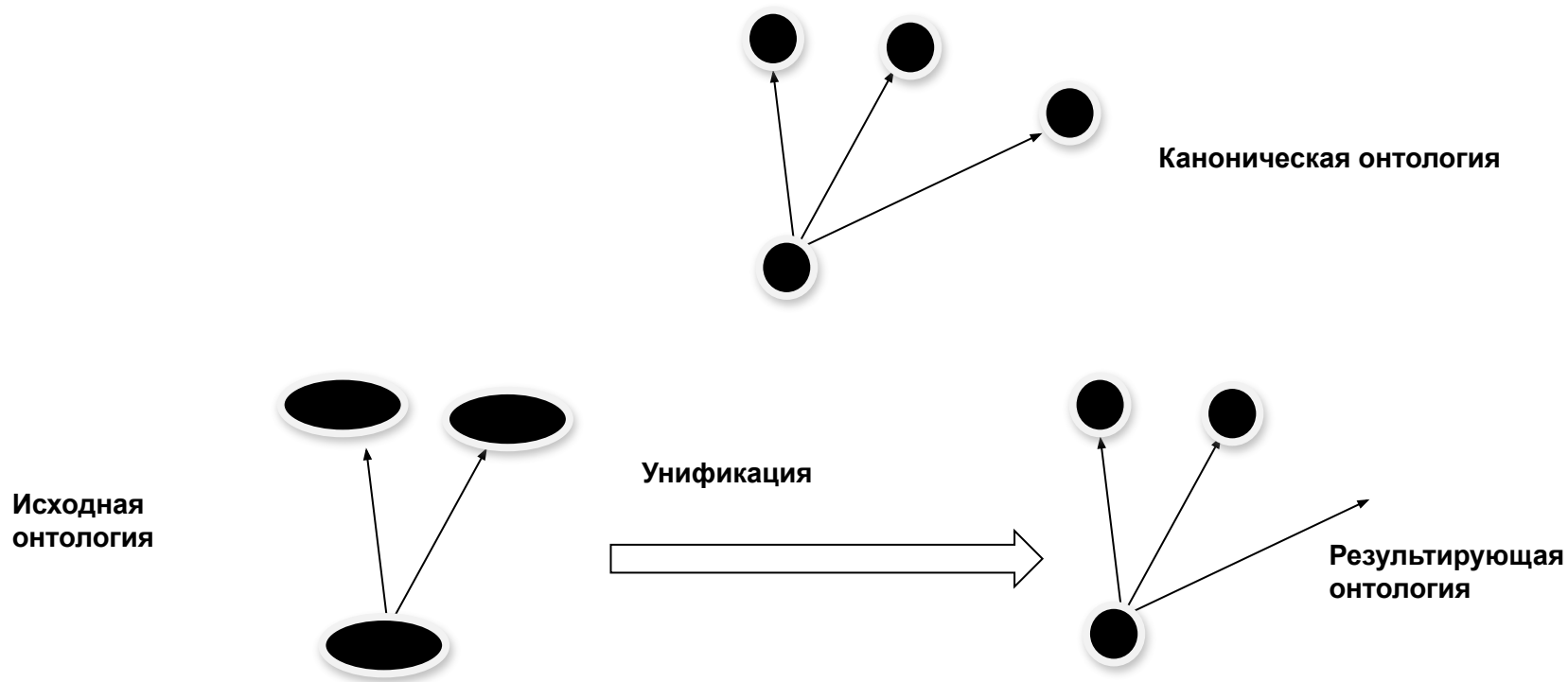
- ❖ Сопоставление (matching)
- ❖ Унификация (unification)
- ❖ Уточнение (refinement)
- ❖ Отображение (mapping)
- ❖ Согласование (alignment)
- ❖ Интеграция (integration)
- ❖ Наследование (inheritance)
- ❖ Композиция (composition)
- ❖ Слияние (merging)
- ❖ Объединение, пересечение, разность (union, intersection, difference)

Манипулирование онтологиями – операция сопоставления

- ❖ <http://www.ontologymatching.org/>
- ❖ **Сопоставление онтологий (ontology matching)** является перспективным направлением в решении проблем семантической неоднородности. Сопоставление призвано решать задачи нахождения соответствия между семантически связанными сущностями двух онтологий. Такое сопоставление может использоваться для решения различных задач, например, слияние онтологий (ontology merging), получение ответов на запросы, трансляция данных, навигация по семантическому вебу. Таким образом, сопоставление онтологий позволяет осуществлять оперирование данными и знаниями, представленными в сопоставляемых онтологиях.
- ❖ (**matching = mapping = alignment - ???**)

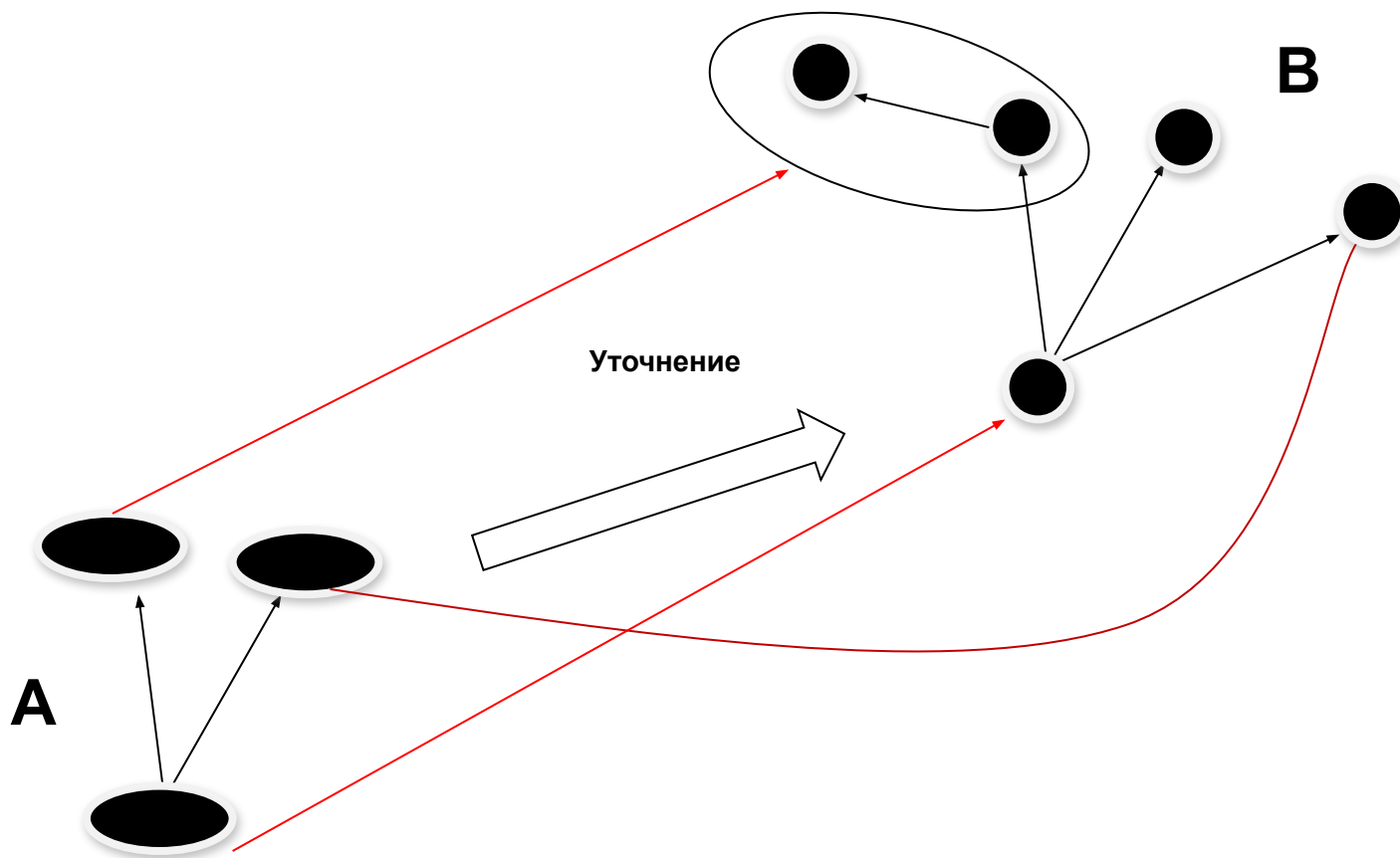
Манипулирование онтологиями – операция унификации

- ❖ **Унификация (unification)** – Онтология приводится к некоему каноническому (эталонному) представлению. Для унификации должна задаваться исходная онтология, которая приводится к результирующей согласно заданной канонической онтологии. Задача унификации множества исходных онтологий становится актуальной при работе с гетерогенными онтологиями.



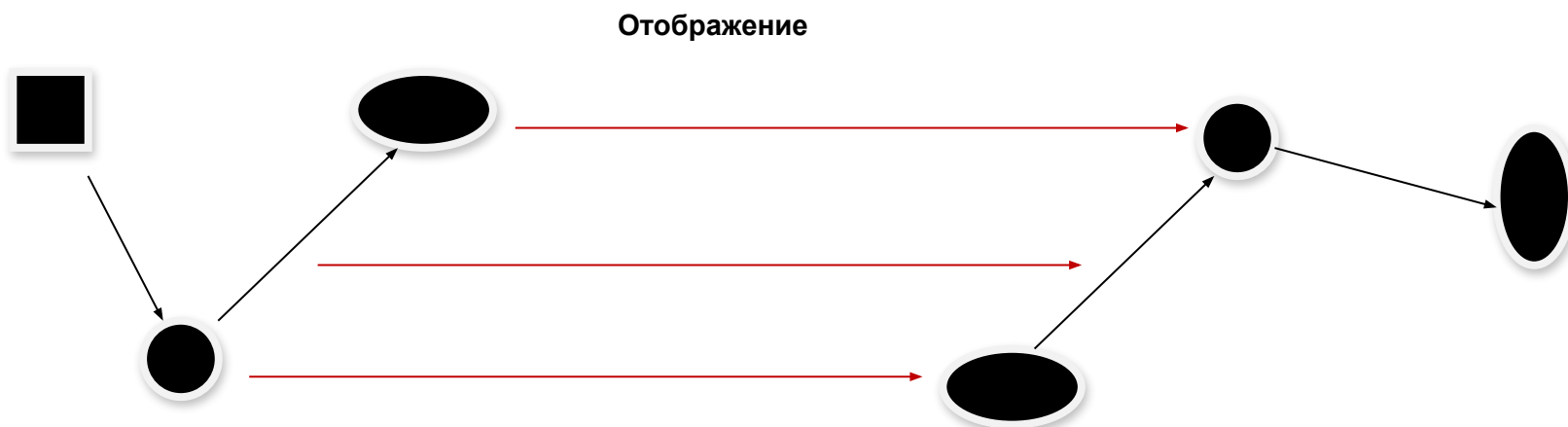
Манипулирование онтологиями – операция уточнения

- ❖ **Уточнение (refinement)**. Под уточнением онтологий понимают такое сопоставление онтологии А с другой онтологией В, что каждому понятию из онтологии А ставится в соответствие эквивалентное ему понятие в В. Прimitives понятия из онтологии А могут соответствовать непримитивным понятиям онтологии В.

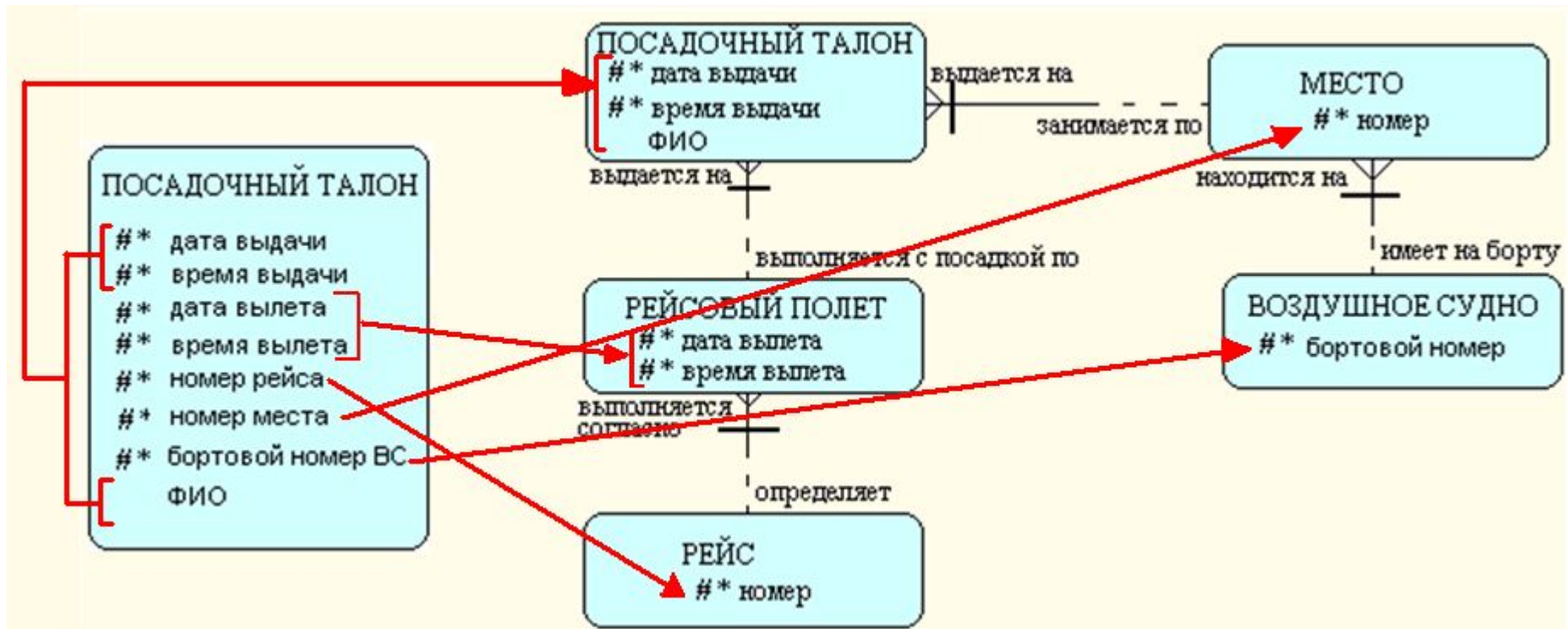


Манипулирование онтологиями – операция отображения

- ❖ **Отображение (mapping).** Отображение одной онтологии в другую – это функция преобразования одной онтологии в другую (способ перевода объектов одной онтологии в другую), либо сам результат такого преобразования. Часто это означает перевод между понятиями и отношениями. Отображение может быть частичным в том смысле, что не все понятия исходной онтологии отображаются в результирующую. В частности, это означает, что в исходной онтологии существует подонтология, для которой существует полное отображение.
- ❖ В результате отображения получаем две онтологии и описания отображения.



Пример отображения



Манипулирование онтологиями – операция согласования

❖ Согласование (alignment) -

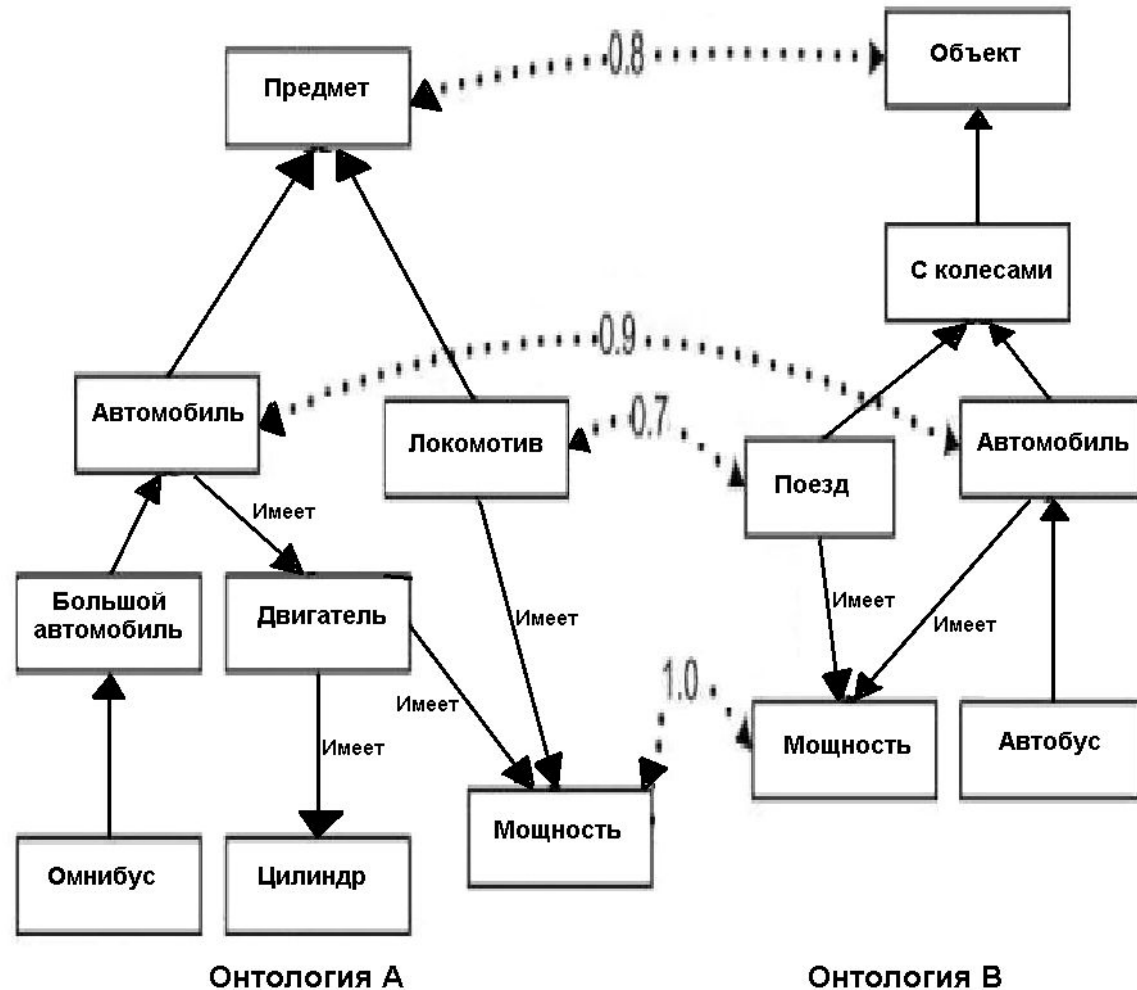
это процесс отображения онтологий в обоих направлениях.

Согласование, как и отображение, может быть лишь частичным.

Спецификация согласования называется **артикуляцией (articulation)**

❖ Согласование заключается в том, чтобы установить

различные виды соответствий между двумя онтологиями, а затем сохранить исходные онтологии вместе с информацией о найденных соответствиях с тем, чтобы в дальнейшем использовать информацию о взаимосвязях онтологий.



Манипулирование онтологиями – операция слияния

- ❖ Слияние (merging) - это процесс создания одной согласованной онтологии из двух или более различных онтологий, относящихся к одной тематической предметной области. Слитая онтология включает информацию из ВСЕХ исходных онтологий с наименьшими возможными изменениями.
- ❖ Слияние может не выполняться полностью, например, в связи с тем, что исходные онтологии могут быть несогласованными (противоречивыми).
- ❖ Результирующая онтология может содержать новые понятия и отношения, которые служат для установления соответствия между терминами и отношениями исходных онтологий.

Манипулирование онтологиями – операции интеграции и наследования

❖ **Интеграция (integration).** Это процесс поиска одинаковых частей (то есть установление соответствия) двух разных онтологий, А и В, при разработке новой онтологии С, которая позволяет выполнить перевод между онтологиями А и В, и, таким образом, позволяет взаимодействие между двумя системами, где одна использует онтологию А, а другая - онтологию В. Новая онтология С может заменить онтологии А и В или может использоваться в качестве промежуточной онтологии для перевода между двумя онтологиями. Интеграция может меняться от согласования к унификации.

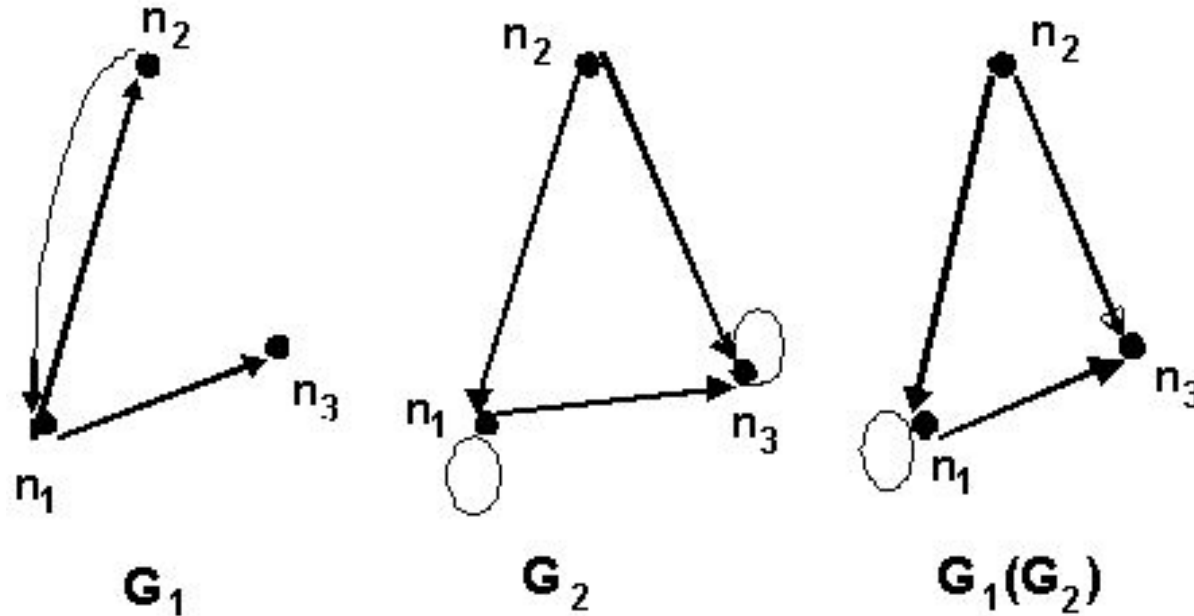
Отличие интеграции от слияния – merging (см. выше) :

- Интеграция устанавливает соответствие, а слияние дает одну результирующую онтологию
- Интеграция, как правило, действует над онтологиями различных доменов, а слияние – одного домена.

❖ **Наследование (inheritance).** Означает, что онтология А наследует все из онтологии В. Она наследует все понятия, отношения и ограничения или аксиомы, и дополнительные знания, содержащиеся в онтологии, не внося при этом какой-либо несогласованности.

Манипулирование онтологиями – операция композиции

Композиция (composition) двух графов G_1 и G_2 , $G=G_1(G_2)$ – это транзитивное замыкание двух графов, если рассматривать их как бинарные отношения.

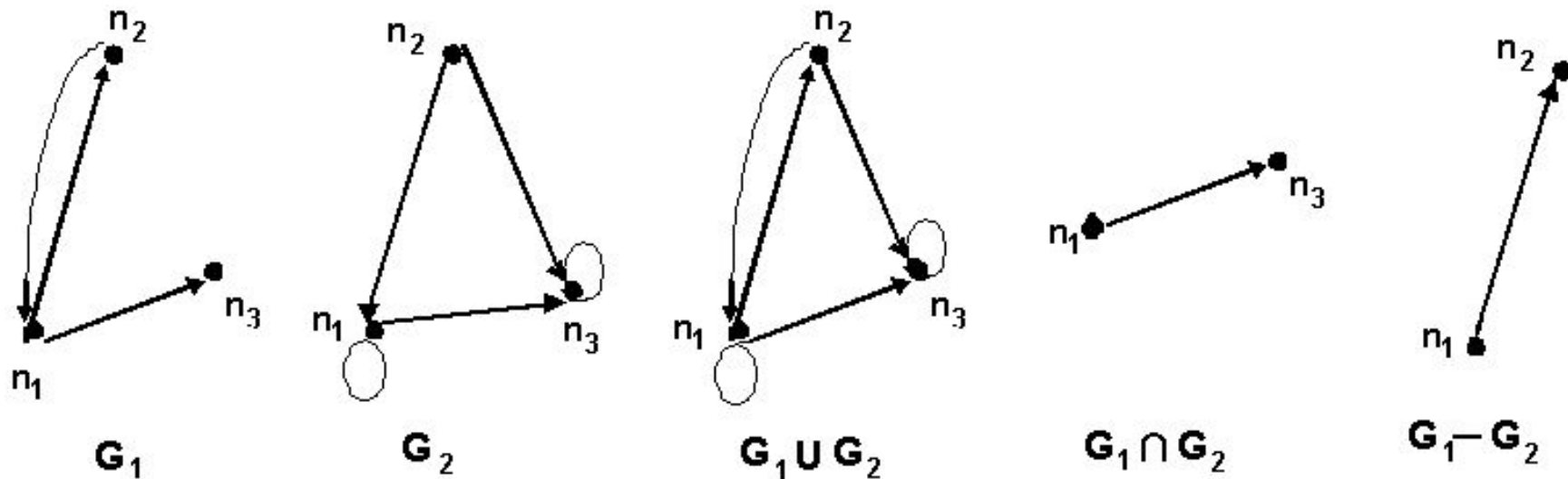


G_1	G_2	$G_1(G_2)$
(n_1, n_2)	(n_2, n_1)	(n_1, n_1)
	(n_2, n_3)	(n_1, n_3)
(n_1, n_3)	(n_3, n_3)	(n_1, n_3)
(n_2, n_1)	(n_1, n_1)	(n_2, n_1)
	(n_1, n_3)	(n_2, n_3)

Дед $(X, Y) = \text{Отец}(\text{Отец}(X, Y))$

Манипулирование онтологиями – операции объединения, пересечения и разности

- ❖ Операции **объединения, пересечения и разности онтологий**. Это обычные теоретико-множественные операции, предполагающие, что онтология – это множество бинарных отношений.



G_1	G_2	$G_1 \cup G_2$	$G_1 \cap G_2$	$G_1 - G_2$
(n_1, n_2)	(n_2, n_1)	n_1, n_2	(n_1, n_3)	(n_1, n_2)
(n_1, n_3)	(n_2, n_3)	(n_1, n_3)		
(n_2, n_1)	(n_3, n_3)	(n_2, n_1)		
	(n_1, n_1)	(n_2, n_3)		
	(n_1, n_3)	(n_3, n_3)		
		(n_1, n_1)		

Вывод (рассуждения) в онтологиях (ontology reasoning)

- ❖ Вывод (inference, reasoning) предполагает возможность выводить новые неявно заданные данные (знания) из имеющихся явно заданных.
- ❖ Решать проблему вывода в онтологии в достаточной степени полноты можно в том случае, когда онтология формализована.
- ❖ Языком формализации онтологий является дескриптивная логика (DL), которая формализует такие два языка онтологий, как RDF и OWL.
- ❖ Используется на большинстве этапов жизненного цикла онтологии.

Способы задания вывода

- ❖ Вывод можно задавать **процедурно**, например, указывая, что определяемое понятие является последовательностью операций над уже определенными понятиями. Например, можно указать, что понятие ДЕД является композицией (транзитивным замыканием) понятия ОТЕЦ.

$$\text{ДЕД}(X, Y) = (\text{ОТЕЦ}(\text{ОТЕЦ}(X, Y)))$$

- ❖ Вывод также можно задавать посредством **правил вывода**. Они могут быть логическими (дедуктивные, индуктивные и т.д.) или прикладными, то есть правилами типа “Если...то”, описывающие семантику предметной области. Например,

ЕСЛИ ОТЕЦ(x, y) & ОТЕЦ(y, z) ТО ДЕД(x, z)

ЕСЛИ СОТРУДНИЧАЕТ(x, y) ТО СОТРУДНИЧАЕТ(y, x)

Вывод в RDF, OWL, DL

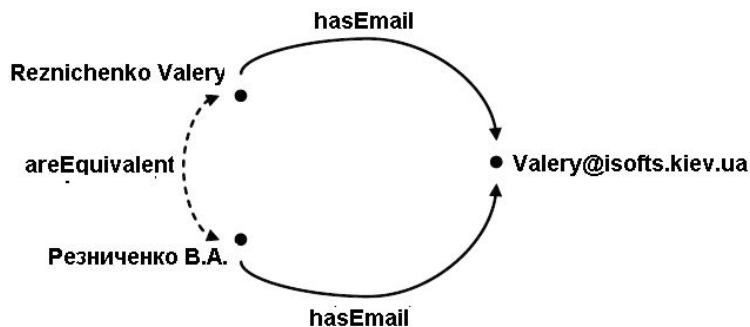
- ❖ В основу RDF, OWL и DL положены такие понятия, как **индивид**, **класс** и **свойство** (свойство вбирает в себя такие традиционные понятия, как атрибут, бинарная связь (отношение), аксиома). В связи с этим, если за основу описания онтологий берутся эти языки, важно иметь выводы, которые характеризуют эти три составляющие онтологии. Поэтому можно говорить о выводах на уровне
 - индивидов,
 - классов,
 - свойств,
 - аксиом

Вывод на индивидах

- ❖ Основная проблема вывода заключается в проверке, допустим ли индивид в ПО. Если предположить, что индивид может существовать только посредством его принадлежности тому или иному классу, то тогда вопрос существования индивида в ПО сводится к проверке его принадлежности классу. Класс, с точки зрения его экстенционала может задаваться:
 - явным перечислением принадлежащих ему индивидов, то есть экстенционально (в этом случае индивид имеется в ПО, если он явно задан приписан классу)
 - указанием характеристик, которым должен обладать индивид, чтобы являться экземпляром класса – интенциональное определение класса.
 - указанием процедуры получения элементов данного класса из элементов других классов.

В связи с этим механизмы проверки принадлежности индивида классу могут быть разными.

- ❖ Еще одной задачей вывода является установление, являются ли два индивида тождественными (эквивалентными) или нет. Для установления этого факта могут использоваться свойства классов, которым они принадлежат, а также их личные свойства. На приведенном ниже рисунке дается графическое представление правила, выраженное на конкретных индивидах: «лица, имеющие один и тот же e-mail, являются одним и тем же лицом», и два способа задания этого правила: в виде правила вывода и введением нового понятия, выводимого из существующих



$$\text{hasEmail}(x, z) \ \& \ \text{hasEmail}(y, u) \ \& \ z = u \ \rightarrow \ x = y$$

$$\text{areEquivalent}(x, y) \equiv$$

$$\forall z \ \forall u (\text{hasEmail}(x, z) \ \& \ \text{hasEmail}(y, u) \ \& \ z = u)$$

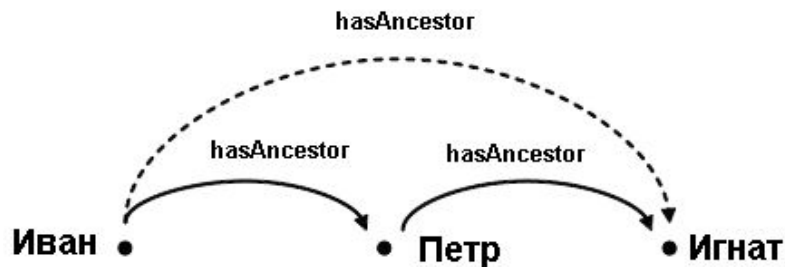
Вывод на классах

- ❖ Основная задача вывода на классах – это, в общем случае, определение взаимоотношения классов между собой, и, в частности, является ли класс В подклассом класса А. Эта частная задача называется задачей **обобщения-специализации** классов (задача выявления **родовидовых зависимостей** на классах). По сути задача обобщения – это выявление факта, являются ли критерии принадлежности одному из классов частным случаем (специализацией, родом, выводимыми из) критериев принадлежности другому классу.
 - Если критерии принадлежности классам взаимовыводимы, то они **эквивалентны**.
 - Если класс принадлежит пустому классу, то его критерии **противоречивы** (невыполнимы).
 - Полный граф обобщения, построенный на всех классах онтологии называется **таксономической иерархией классов**.
- ❖ Существует также задачи вывода на классах, связанные с установлением и формализацией других видов взаимоотношения между классами (экземпляр-класс, часть-целое, агрегат-деталь)

Вывод на свойствах

- ❖ Вывод на свойствах – это вывод новых свойств из существующих.
- ❖ Так как свойства – это бинарные отношения, то в качестве вывода могут использоваться традиционные свойства бинарных отношений: рефлексивность, симметричность, транзитивность, а также их отрицания: иррефлексивность, ассиметричность и антитранзитивность.

Пример использования транзитивности свойства Предок (Ancestor):



- ❖ Так как бинарные отношения являются множествами, то можно говорить об **обобщении/специализации свойств**. Например,
 - свойство **Иметь ребенка** является обобщением свойств **Иметь сына**, **Иметь дочь**,
 - свойства **Являться младшим (старшим, ведущим,...) научным сотрудником** является специализацией свойства **Являться научным сотрудником**.
- ❖ В связи с этим для свойств существуют такие же проблемы вывода, как и для классов (обобщение/специализация, эквивалентность, выполнимость, таксономия свойств).

Вывод на аксиомах

- ❖ Является ли онтология (система аксиом) полной?
- ❖ Является ли онтология (система аксиом) непротиворечивой?
- ❖ Является ли то или иное понятие (сущность, атрибут, связь) или подонтология выводимой из имеющейся?

Жизненный цикл - этап развития



- ❖ **Сбор предложений** по развитию (изменению)
- ❖ **Представление предложений** по развитию - чтобы реализовать требования по изменению онтологии, они должны быть представлены в подходящей форме и с требуемой степенью детализации (простые изменения, сложные изменения)
- ❖ **Семантика развития/изменений** - изменение/развитие онтологии должно производиться с учетом выработанных на этапе анализа требований по ее целостности. Целостность, это, по-сути, аксиоматика онтологии.
- ❖ **Реализация изменений**
- ❖ **Распространение изменений** - изменяемая онтология может быть связана тем или иным образом (mapping, alligning) с другими онтологиями. Поэтому ее изменения должны быть распространены на другие связанные с ней онтологии.
- ❖ **Проверка** – тестирование правильности проведенных изменений

Ontology learning

- ❖ **OntoWeb Consortium. Deliverable 1.5: A survey of ontology learning methods and techniques. – 86 p.**
- ❖ **Ontology learning (познание) (ontology extraction, ontology generation, or ontology acquisition)** – это полуавтоматическое извлечение необходимых понятий и отношений между ними из соответствующих (неструктурированных, полуструктурированных и структурированных) источников с целью построения онтологии. Это сложное междисциплинарное направление исследований, которое использует знания из таких дисциплин, как обработка естественных языков, интеллектуальный анализ данных (data mining), машинное обучение и представление знаний.

Источники для **Ontology learning**

- ❖ **Ontology learning на основании тестов** – извлечение онтологий посредством применения к текстам методов анализа естественных языков
- ❖ **Ontology learning на основании словарей** – использование машиночитаемых словарей для извлечения из них понятий и связей между ними
- ❖ **Ontology learning на основании баз знаний** - построение онтологий на основании сведений, имеющих в базах знаний
- ❖ **Ontology learning из полуструктурированных источников** – построение онтологий на основании источников, имеющих определенную структуру, например, схемы XML
- ❖ **Ontology learning из реляционных схем** – построение онтологий на основании понятий и отношений между ними, извлеченных их баз данных

«Слоеный пирог» Ontology learning

$\forall x \forall y (\text{страдает-от}(x, y) \rightarrow \text{болеет}(x))$

лечение (домен:доктор домен:болезнь)

is_a (доктор, человек)

болезнь:=<I, E>

{болезнь, заболевание}

болезнь, заболевание, больница

Аксиомы и правила

Отношения

Таксономия

Понятия

Синонимы

Термины

Инженерия онтологий - языки онтологий

