



СПбГЭТУ «ЛЭТИ»

Кафедра Вычислительной техники

Магистерская программа

«Семантические технологии и многоагентные системы»

Дисциплина «Семантический Web»

Лекция 3

Онтологии:

Классификации & примеры

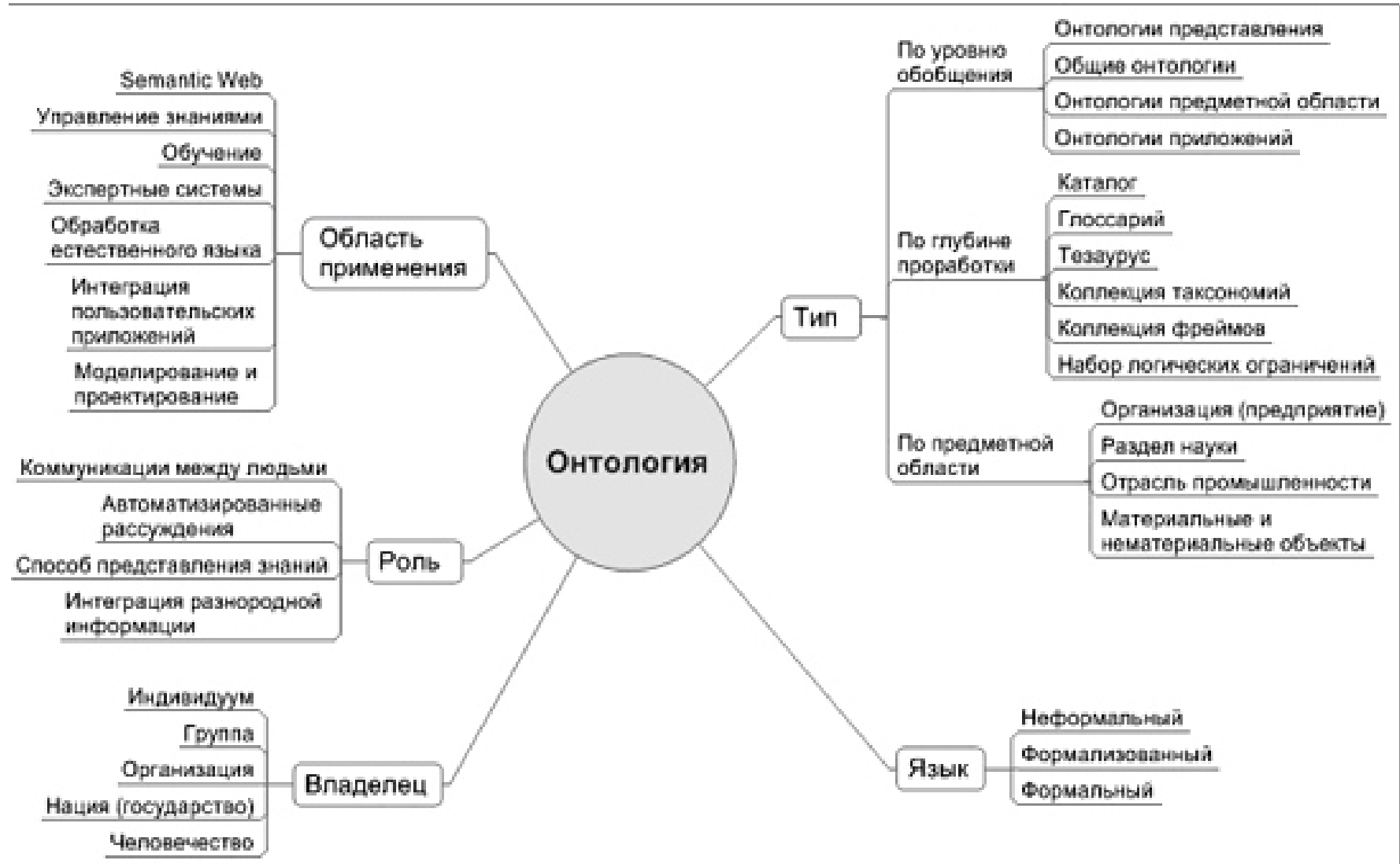
Содержание лекции

- Определение онтологий
- Классификации онтологий
- Примеры существующих онтологий
- Требования к онтологиям
- Методологии проектирования онтологий
- Согласование онтологий

Определение онтологий

- В философии:
 - *Онтология* – учение о сущем, о бытии как таковом;
 - изучает фундаментальные принципы бытия, его наиболее общие сущности и категории, принципы, структуры и закономерности
 - в противоположность *гносеологии* - теории познания
- В информатике:
 - *Онтология* – явная (эксплицитная, формальная) спецификация концептуализации
 - *Концептуализация* – описание на уровне понятий множества сущностей (классов, объектов, отношений) некоторой предметной области

Классификация онтологий



Два подхода к построению онтологий

- Исторически сложились **два подхода** к построению и исследованию онтологий:
 - **Формальный** – основан на формальных логиках (предикатов первого порядка, *дескриптивной*, модальной и др.);
 - **Лингвистический** – основан на изучении естественного языка (в частности, его семантики) и построении онтологий на больших текстовых массивах
- Эти подходы взаимосвязаны:
 - *лексические онтологии* содержат элементы формальных аксиоматик;
 - логические системы включают лингвистические знания;
- Согласно одному из первых подходов к классификации онтологий их делят по 3 признакам (*Lassila & McGuinness, 2001*) :
 - *степень формальности*;
 - *цель создания*;
 - *содержание*

Классификация онтологий по степени формальности



Формальное определение онтологии (базовое): $O = \langle C, R, I \rangle$, где

- C – множество концептов (терминов);
- R – множество отношений между концептами (ролей);
- I – формальная интерпретация.

Степени формализации онтологий: *Управляемые словари и глоссарии*

Каталоги на основе идентификаторов (управляемые словари – controlled vocabulary):

- конечный список терминов, имеющих ID;
- представляют **точную** (однозначную) **интерпретацию** терминов;
- например, термину "машина" при любом его использовании будет соответствовать значение 25



Глоссарии (словари терминов):

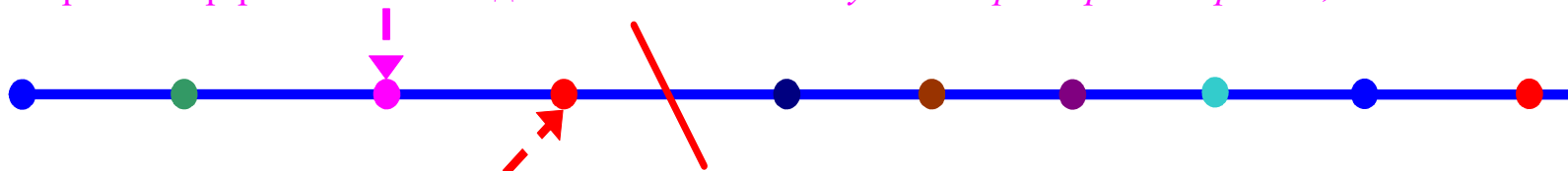
- список терминов *с их значениями*;
- значения *описываются* (определяются) *неформально* в виде комментариев на естественном языке, ориентированных на людей;
- дают больше информации, поскольку люди могут понять смысл термина;
- допускают многозначные интерпретации терминов;
- *непригодны для автоматической обработки программными агентами*;
- терминам *могут быть* присвоены ID.

Степени формализации онтологий: Тезаурусы и неформальные таксономии

• **Тезаурусы** – словари, содержащие *семантические отношения* между терминами:

- *синонимы* – ...;
- *антонимы* – ...;
- *паронимы* – частичное звуковое сходство слов при их полном или частичном семантическом различии: невежа/невежда, орудие/оружие;
- *гипонимы* – слова с более узким значением, называющие предмет (свойство, признак) как элемент класса (множества)
- *гиперонимы* – слова с более широким значением, выражающие общее, родовое понятие, название класса (множества) предметов (свойств, признаков)

- в некоторых случаях эти отношения могут однозначно интерпретироваться агентами;
- обычно тезаурусы не предоставляют явной иерархии;
- простые иерархии можно выводить с использованием *сужения и расширения терминов*;



• **Неформальные таксономии:**

- вводится (неформально !) отношение *isA*;
- свойство транзитивности отношения *isA* может не выполняться;
- Например: Yahoo ввела категорию верхнего уровня - "предметы одежды". Затем подкатеорию "женские". Затем подкатегории "аксессуары" и "платья". Аксессуары (например "броши"), не являются предметами одежды

Степени формализации онтологий: *Формальные таксономии и экземпляры*

- **Формальные таксономии:**
 - онтологии, включающие *строгое определение* отношения «Класс-Подкласс» - isA ;
 - для классов *строго соблюдается транзитивность* отношения isA :
 - если B – подкласс класса A , то каждый подкласс класса B также является подклассом класса A ;
 - *строгое определение иерархий подклассов позволяет использовать наследование в процедурах логического вывода.*
- **Формальные экземпляры:**
 - *строго определяется* отношение «Класс-Экземпляр» ($isInstanceOf$):
 - если B – подкласс класса A , то каждый экземпляр класса B также является экземпляром класса A ;
 - некоторые классификации включают только имена классов, другие содержат на нижнем уровне экземпляры (*индивиды*).

Степени формализации онтологий: *Фреймы и ограничения на значения*

- **Фреймы:**

- Классы (фреймы) включают *информацию о свойствах*;
 - Например, класс «Предмет одежды» может иметь свойства «цена» и «сделанИз»;
- Свойства во фреймовых системах *определяются на верхних уровнях иерархии и наследуются подклассами*;
 - Например, в онтологии коммерции класс «Товар» может иметь свойство "цена", которое наследуется всеми его подклассами.

- **Ограничения на значения:**

- Возможность накладывать *ограничения на значения свойств*;
 - Например, значения свойства “цена” могут быть ограничены *числовыми значениями* определенного *диапазона*, свойства «сделанИз» - экземплярами класса «Материал» и т.д.
- В классификационных схемах, *не поддерживающих строгое отношение «is-a», могут возникать проблемы*:
 - Если класс «Духи» определены как подкласс класса «Предмет одежды», он унаследует свойство "сделанИз" вместе с ограничением ("Материал").

Степени формализации онтологий: *Общие логические ограничения*

- ***Общие логические ограничения:***
 - выразительные средства для описания в онтологии более сложных фактов;
 - определение некоторого свойства экземпляра с использованием математического выражения на основе значений других свойств данного экземпляра или значений свойств других экземпляров:
 - Масса тела равна произведению его плотности на объем;
 - возможность делать произвольные логические утверждения (Ontolingua, СуsL, ...);
 - специфицировать *ограничения между термами с помощью логики первого порядка*;
 - объявлять два и более класса непересекающимися (не имеющими общих экземпляров);
 - объявлять инверсные отношения;
 - объявлять отношение «Часть-целое»;
 - . . .

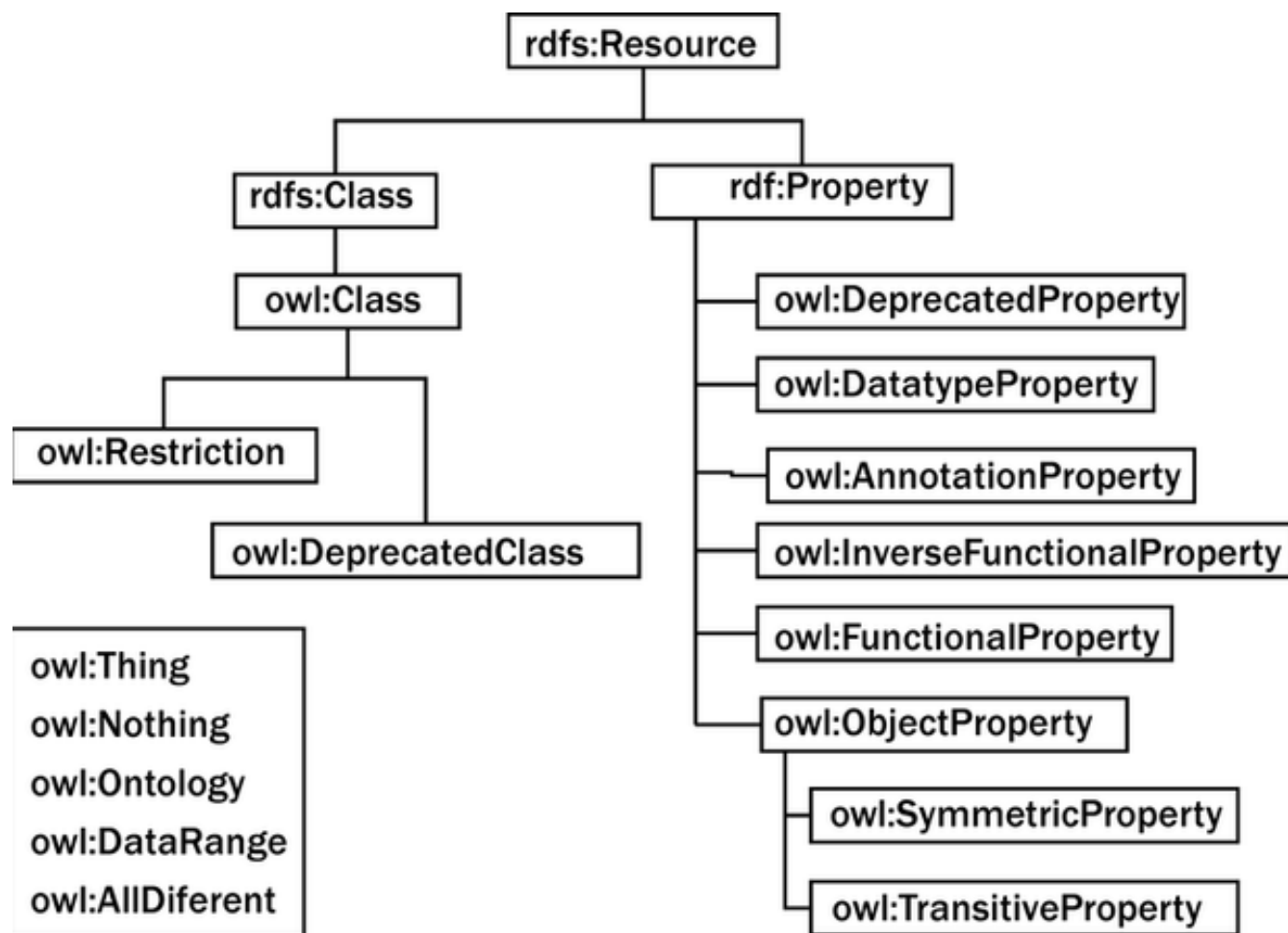
Классификация онтологий *по цели создания*

- По *цели создания* выделяют 4 типа онтологий:
 - онтологии *представления*;
 - онтологии *верхнего уровня* (top-level ontology);
 - онтологии *предметных областей* (domain ontology);
 - *онтологии приложений* (application ontology).

Онтологии представления

- *Онтологии представления:*
 - Предназначены для концептуализации *формализмов представления знаний*;
 - *Определяют язык* для спецификации других онтологий более низких уровней
 - Пример: описание понятий языка OWL средствами RDF/RDFS
 - По сути являются *метаонтологиями*

Онтология представления для языка OWL



Онтологии верхнего уровня

- Предназначены для многократного использования в *разных предметных областях*
- Фиксируют **общие для разных предметных областей знания**
- Содержат **общие концепты**: «Сущность», «Явление», «Процесс», «Объект», «Роль» и т. п.
- Онтологии верхнего уровня *похожи друг на друга*;
- Известные проекты:
 - The Standard Upper Merged Ontology (SUMO) (<http://suo.ieee.org/>)
 - **Сyc'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>)
- Попытки создать единую онтологию верхнего уровня **пока успехом не увенчались**

Онтология верхнего уровня SUMO

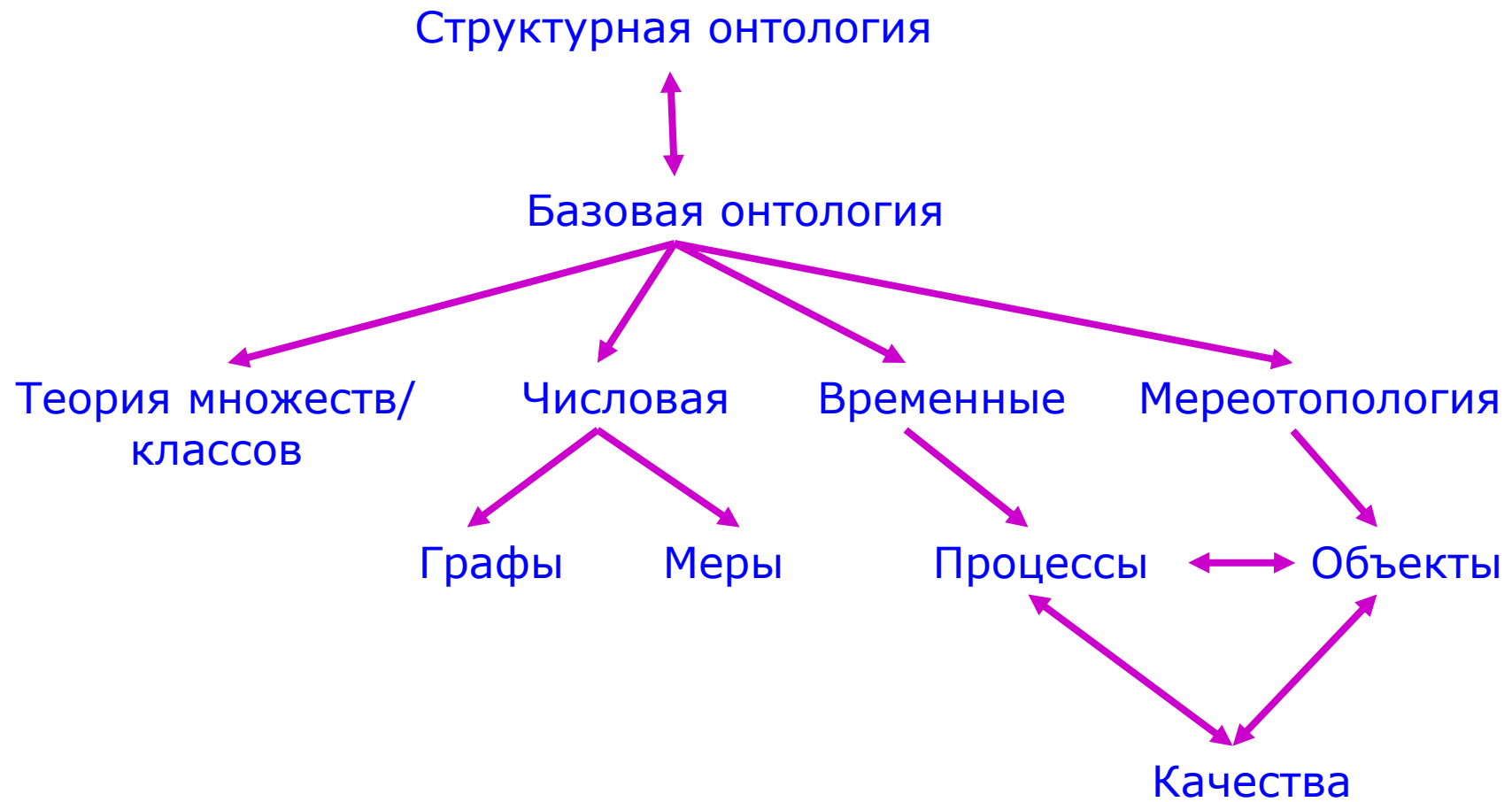
<http://www.ontologyportal.org/>

- **SUMO (Suggested Upper Merged Ontology):**
 - крупнейшая доступная формальная онтология;
 - разработана в рамках проекта рабочей группы IEEE SUO (IEEE Standard Upper Ontology Working Group) и фирмы Teknowledge (<http://www.adampease.org/OP/>);
 - претендует на статус стандарта для онтологий верхнего уровня;
 - содержит *наиболее общие и абстрактные* концепты;
 - содержит **~25 000** понятий и **~80 000** аксиом, определяющих эти понятия;
 - назначение – содействовать улучшению интероперабельности данных, извлечения и поиска информации, автоматического вывода (доказательства), обработки естественного языка;

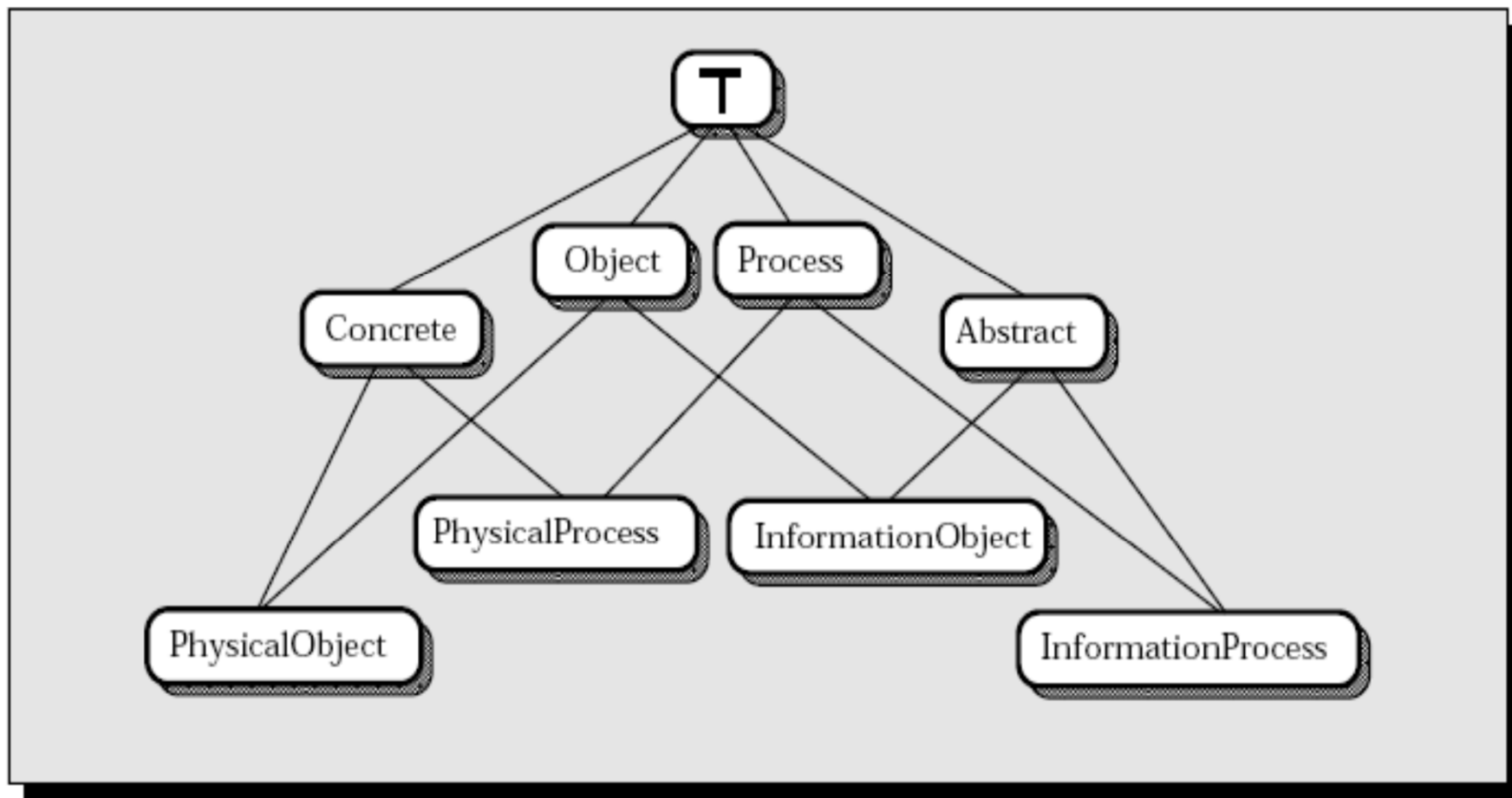
Онтология верхнего уровня SUMO

- охватывает следующие типы сущностей:
 - общие виды процессов и объектов,
 - абстракции (теория множеств, атрибуты, отношения),
 - числа и единицы измерения,
 - темпоральные (временные) понятия,
 - части и целое,
 - агенты и намерения.
- является «канонической» онтологией верхнего уровня:
 - содержит обозримое число концептов и аксиом,
 - имеет ясную иерархию классов,
 - легко расширяется,
 - является итогом объединения различных общедоступных онтологий верхнего уровня (в том числе онтологии *Джона Совы*)
 - поддерживает возможность трансляции описания онтологии на любой из основных языков представления знаний,
 - наличие онтологии среднего уровня (MIL0), гладко интегрированной с SUMO,
 - множество примеров практического применения,
 - **связь с WordNet** – наиболее крупным на настоящий момент тезаурусом, содержащим около 150 тыс. слов повседневного английского языка.

Фрагмент онтологии верхнего уровня SUMO (Верхний уровень)



Фрагмент онтологии верхнего уровня Дж.Совы (John Sowa)



Онтология верхнего уровня DOLCE

- **DOLCE** (Descriptive Ontology for Linguistic and Cognitive Engineering)
 - разработана в рамках **проекта WonderWeb** (<http://www.loa.istc.cnr.it/dolce/overview.html>), является первой в библиотеке базовых онтологий;
 - не претендует на звание универсальной, стандартной или общей;
 - ориентирована на **применение в Semantic Web** для *согласования между интеллектуальными агентами*, использующими разную терминологию;
 - Цель: *создать модель, помогающую при сравнении и объяснении связей с другими онтологиями библиотеки WFOL* (базовой библиотеки онтологий WonderWeb), а также для *выявления скрытых допущений*, лежащих в основе существующих онтологий и лингвистических ресурсов, таких как WordNet;
 - имеет *когнитивный уклон*, поскольку фиксирует *онтологические категории естественного языка* и знания «здравого смысла».

Онтология верхнего уровня DOLCE

- В основу процесса проектирования положено фундаментальное философское разделение всех сущностей на *универсалии* (сущности потенциально или реально имеющие экземпляры) и *индивиды* (или частности), которые не имеют и не могут иметь экземпляров.
- DOLCE - *онтология индивидов*, область описания ограничена только ими
 - Пример универсалии - понятие «Собака» - имеет множество экземпляров, конкретных примеров в окружающем мире.
 - В отличие от этого понятия, понятие «Время» скорее рассматривается как индивид.
- Явное разделение на «Постоянные» и «Происходящие» сущности.
 - «Постоянные» сущности *имеются в наличии целиком и неизменно* в некотором фиксированном промежутке времени – в течение периода своего существования. Например, «Стол», «Дом».

Фрагмент онтологии верхнего уровня DOLCE



Онтологии предметных областей

- Онтологии *предметных областей*:
 - предназначены для повторного использования знаний *о конкретной предметной области*:
 - образование,
 - медицина,
 - культура,
 - вычислительная техника и т. д.;
 - круг решаемых задач и вопросов, на которые онтология отвечает, *ограничен выбранной областью*;
 - в свою очередь иерархически декомпозируются на подобласти;
 - **большой** по сравнению с ОВУ объем онтологии;
 - наличие отношений специфичных для конкретной области.
- Повторное использование возможно только в рамках данной предметной области

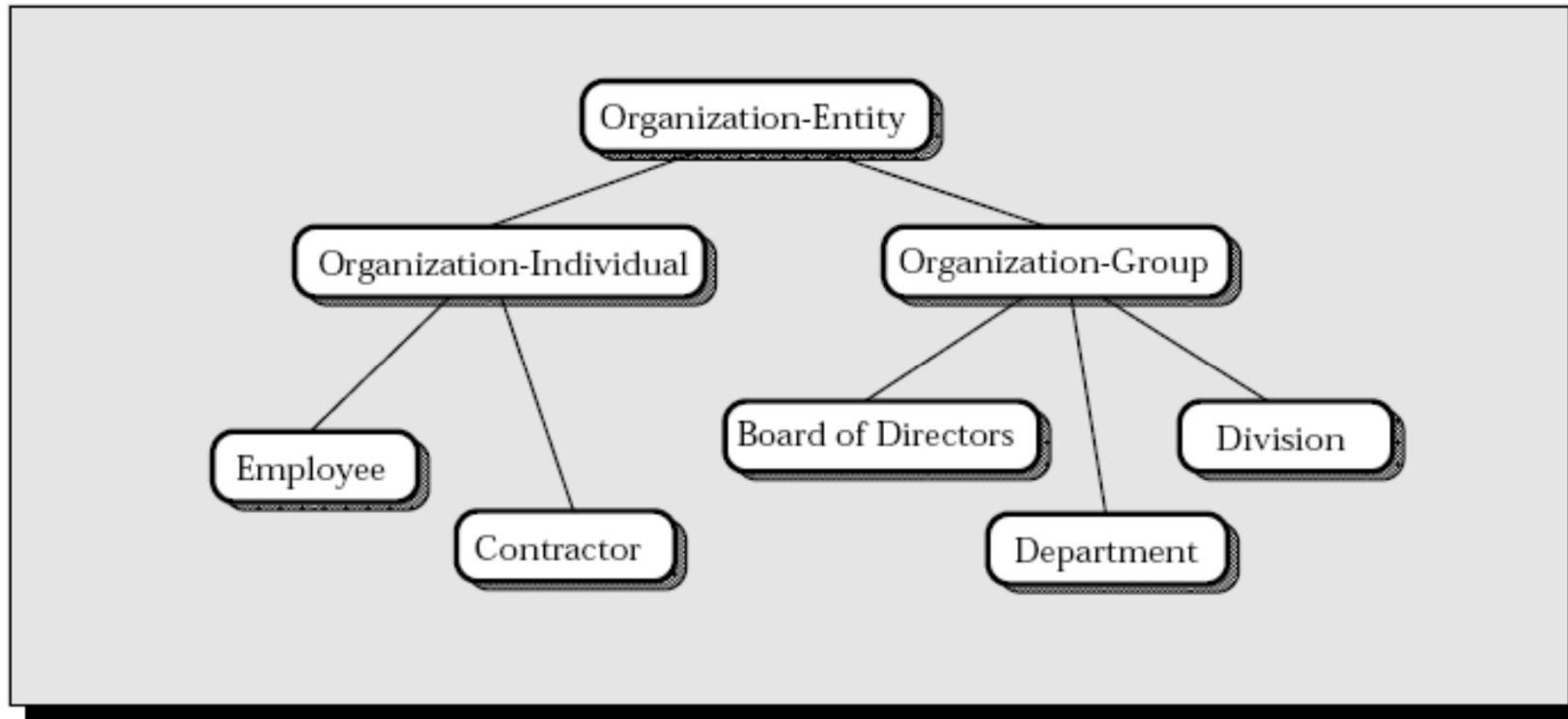
Онтологии предметных областей

- Онтологии предметной области *Электронная коммерция*:
 - The United Nations Standard Products and Services Codes (UNSPSC)
(<http://www.unspsc.org/>)
 - NAICS (North American Industry Classification System)
(<http://www.census.gov/epcd/www/naics.html>)
 - SCTG (Standard Classification of Transported Goods)
(<http://www.statcan.ca/english/Subjects/Standard/sctg/sctg-menu.htm>)
 - E-cl@ss (<http://www.eclass.de/>)
 - RosettaNet (<http://www.rosettanet.org>)

Онтологии приложений

- Онтологии приложений:
 - описывают концептуальную модель *конкретной задачи* или *приложения*;
 - содержат наиболее специфичную информацию;
 - в меньшей степени ориентированы на повторное использование;
 - Примеры: TOVE, Plinius. . .
 - TOVE – проект по разработке множества интегрированных онтологий для *моделирования предприятий*
- <http://www.eil.utoronto.ca/enterprise-modelling/tove/>

Фрагмент онтологии TOVE



Классификация онтологий *по содержанию*

- По *содержимому* выделяют 3 типа онтологий:
 - *общие* онтологии – соответствуют онтологиям верхнего уровня;
 - онтологии *задач* – соответствуют задачам решаемым в проблемной области (составление расписаний, классификация, управление. . .);
 - *предметные* онтологии - соответствуют онтологиям предметных областей.

Лексические (лингвистические) онтологии

- Фиксируются совместно *понятия (термины) и их языковые свойства*
 - Понятиям онтологии сопоставляются наборы языковых конструкций (слов и словосочетаний), которыми они могут выражаться в тексте (в разных языках по-разному!)
- Начали разрабатываться в целях *автоматизации обработки текстов на естественном языке* (Natural Language Processing, NLP) и *информационного поиска*
- Основные отношения в лингвистических онтологиях:
 - синонимия;
 - гипонимия (выражение подвида);
 - меронимия (выражение составной части);
- Наиболее известные *лингвистические онтологии*:
 - **WordNet** (<http://www.globalwordnet.org/>);
 - EuroWordNet (<http://www.hum.uva.nl/~ewn/>);
 - **Русский WordNet** (<http://wordnet.ru/>);
 - The Generalized Upper Model (<http://www.darmstadt.gmd.de/publish/komet/genum/newUM.html>);
 - The Mikrokosmos ontology (<http://crl.nmsu.edu/mikro> [требуется авторизация]);
 - SENSUS (<http://www.isi.edu/natural-language/projects/ONTOLOGIES.html>);