

# Natural Language Processing

## Лекция 8

Представление знаний для  
компьютерной обработки - часть 1

# Смыслы слов

Многие слова являются **неоднозначными** - одно и то же слово может обозначать несколько вещей.

## Мышь (значения)

---

Материал из Википедии — свободной энциклопедии

**Мышь** может обозначать:

- **Мыши** (лат. *Muridae*) — семейство млекопитающих отряда грызунов.
- **Домовая мышь** (лат. *Mus musculus*) — вид грызунов рода домовых мышей.
- **Летучие мыши** (лат. *Microchiroptera*) — семейство млекопитающих отряда рукокрылых.
- **Птицы-мыши** (лат. *Coliiformes, Coliidae*) — отряд птиц с единственным одноимённым семейством.
- **Компьютерная мышь** — одно из устройств ввода информации в компьютер.
- **«Мышь»** — сверхтяжёлый немецкий танк.
- **Пол-литровая мышь** — персонаж одноимённого мультфильма.
- **Мышь (галактика)** — классический образец столкновения **галактик**.
- **Мышь (рыбалка)** — искусственная рыболовная приманка.
- **Мышь** — манга и аниме в жанре этти.
- **«Мыши»** — 15-я серия 1-го сезона телесериала «За гранью возможного».

# СМЫСЛЫ СЛОВ

**Смысл** (смысл слова, word sense) - это отдельное представление одного аспекта.

**Тезаурус** - особая разновидность словарей, в которых указаны семантические отношения (синонимы, антонимы, паронимы, гипонимы, гиперонимы и т. п.) между лексическими единицами.

**Разрешение лексической многозначности** (word sense disambiguation, WSD) — это проблема обработки естественного языка, которая заключается в задаче выбора значения (или смысла) многозначного слова или словосочетания в зависимости от контекста, в котором оно находится.

# Определение смысла слова

- Векторные представления (word2vec, ELMo, BERT и др.);
- Словари;
- Тезаурусы

right *adj.* located nearer the right hand esp. being on the right when facing the same direction as the observer.

left *adj.* located nearer to this side of the body than the right.

red *n.* the color of blood or a ruby.

blood *n.* the red liquid that circulates in the heart, arteries and veins of animals.

# Сколько значений имеет слово?

Какие необходимые критерии нужны для того, чтобы у слова выделить отдельные значения?

1. различное синтаксическое поведение;
2. независимые смысловые отношения;
3. они демонстрируют антагонистические значения.

- (1) They rarely **serve** red meat, preferring to prepare seafood.
- (2) He **served** as U.S. ambassador to Norway in 1976 and 1977.
- (3) He might have **served** his time, come out and led an upstanding life.

# Сколько значений имеет слово?

**Зевгма** (zeugma) - техника, позволяющая определить, являются ли два смысла одного слова различными или нет.

- (1) Which of those flights *serve breakfast*?
- (2) Does Air France *serve Philadelphia*?
- (3) ? Does Air France *serve breakfast and Philadelphia*?

# Отношения между смыслами: синонимия

**Синонимы** - это слова одной части речи с полным или частичным совпадением значения.

(1.1) How **big** is that plane?

(1.2) How **large** is that plane?

(2.1) Miss Nelson, for instance, became a kind of **big** sister to Benjamin.

(2.2) ?Miss Nelson, for instance, became a kind of **large** sister to Benjamin.

# Отношения между смыслами: антонимия

**Антонимы** - это слова с противоположным значением.

Антонимы полностью различаются по одному аспекту их значения - их положение на шкале или направление, но в остальных аспектах значения они очень похожи. Таким образом, автоматически отличить синонимы от антонимов бывает сложно.



# Отношения между смыслами: таксономия

Слово (или значение) является **гипонимом** другого слова или значения, если первое является более специфичным, обозначает подкласс второго.

Значение **A** является гипонимом значения **B**, если всё, что является **A**, является также и **B**, и следовательно, быть **A** обозначает быть **B**:  $\forall x A(x) \Rightarrow B(x)$ .

Гипонимия/гиперонимия - транзитивное отношение: если **A** - гипоним **B**, и **B** - гипоним **C**, то **A** - гипоним **C**.

Другое название этого отношения - **IS-A**.

# Отношения между смыслами: меронимия

**Меронимия** - отношение часть-целое.

Ножка - **часть** стула; колесо - **часть** машины.

Колесо - **мероним** машины.

Машина - **холоним** колеса.

# Отношения между смыслами: структурная полисемия

Значения слова могут быть также связаны семантически, в этом случае говорят, что их связывает отношение **структурной полисемии**.

Например: Банк - это организация и здание.

**Метонимия** — вид тропа, словосочетание, в котором одно слово заменяется другим, обозначающим предмет (явление), находящийся в той или иной (пространственной, временной и т. п.) связи с предметом, который обозначается заменяемым словом. Замещающее слово при этом употребляется в переносном значении.

AUTHOR ( <i>Jane Austen wrote Emma</i> )	↔	WORKS OF AUTHOR ( <i>I really love Jane Austen</i> )
FRUITTREE ( <i>Plums have beautiful blossoms</i> )	↔	FRUIT ( <i>I ate a preserved plum yesterday</i> )

# Представление знаний

**Представление знаний** (knowledge representation) - это область искусственного интеллекта, изучающая представление информации о мире в форме, пригодной для компьютерной обработки.

Способы представления знаний:

- логические модели;
- семантические сети;
- фреймы;
- онтологии

# WordNet

<http://wordnet.princeton.edu/>

**WordNet** — это лексическая база данных английского языка, разработанная в Принстонском университете. Представляет собой электронный словарь-тезаурус и набор семантических сетей для английского языка.

Слова группируются в наборы синонимов (синсеты), каждое слово снабжено коротким определением и примером использования, а также семантическими отношениями с другими словами.

## WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

### Noun

- [S:](#) (n) **wordnet** (any of the machine-readable lexical databases modeled after the Princeton WordNet)
- [S:](#) (n) **WordNet**, [Princeton WordNet](#) (a machine-readable lexical database organized by meanings; developed at Princeton University)

# WordNet

WordNet содержит 155,827 слов, организованных в 117,659 синсетов (206,941 пара слово-значение).

WordNet содержит существительные, глаголы, наречия и прилагательные.

# WordNet

- S: (adj) dependable, **good**, safe, secure (financially safe) *"a good investment"; "a secure investment"*
- S: (adj) **good**, right, ripe (most suitable or right for a particular purpose) *"a good time to plant tomatoes"; "the right time to act"; "the time is ripe for great sociological changes"*
- S: (adj) **good**, well (resulting favorably) *"it's a good thing that I wasn't there"; "it is good that you stayed"; "it is well that no one saw you"; "all's well that ends well"*

*Examples of adjective synsets*



# Thesaurus: WordNet

Semantic relations for nouns:

- **hypernyms:**  $Y$  is a hypernym of  $X$  if every  $X$  is a (kind of)  $Y$  (*canine* is a hypernym of *dog*);
- **hyponyms:**  $Y$  is a hyponym of  $X$  if every  $Y$  is a (kind of)  $X$  (*dog* is a hyponym of *canine*);
- **coordinate terms:**  $Y$  is a coordinate term of  $X$  if  $X$  and  $Y$  share a hypernym (*wolf* is a coordinate term of *dog*, and *dog* is a coordinate term of *wolf*);
- **meronym:**  $Y$  is a meronym of  $X$  if  $Y$  is a part of  $X$  (*window* is a meronym of *building*);
- **holonym:**  $Y$  is a holonym of  $X$  if  $X$  is a part of  $Y$  (*building* is a holonym of *window*).

# Thesaurus: WordNet

- S: (n) **apple** (fruit with red or yellow or green skin and sweet to tart crisp whitish flesh)
  - direct hyponym / full hyponym
    - S: (n) crab apple, crabapple (small sour apple; suitable for preserving)  
*"crabapples make a tangy jelly"*
    - S: (n) eating apple, dessert apple (an apple used primarily for eating raw without cooking)
    - S: (n) cooking apple (an apple used primarily in cooking for pies and applesauce etc)
  - direct hypernym / inherited hypernym / sister term
    - S: (n) edible fruit (edible reproductive body of a seed plant especially one having sweet flesh)
    - S: (n) pome, false fruit (a fleshy fruit (apple or pear or related fruits) having seed chambers and an outer fleshy part)
  - part holonym
    - S: (n) **apple**, orchard apple tree, Malus pumila (native Eurasian tree widely cultivated in many varieties for its firm rounded edible fruits)

# Thesaurus: WordNet

Semantic relations for verbs:

- **hypernym**: the verb *Y* is a hypernym of the verb *X* if the activity *X* is a (kind of) *Y* (*to perceive* is an hypernym of *to listen*);
- **troponym**: the verb *Y* is a troponym of the verb *X* if the activity *Y* is doing *X* in some manner (*to lisp* is a troponym of *to talk*);
- **entailment**: the verb *Y* is entailed by *X* if by doing *X* you must be doing *Y* (*to sleep* is entailed by *to snore*);
- **coordinate terms**: those verbs sharing a common hypernym (*to lisp* and *to yell*).

# Thesaurus: WordNet

- S: (v) **implement** (apply in a manner consistent with its purpose or design)  
*"implement a procedure"*
  - direct hypernym / inherited hypernym / sister term
    - S: (v) use, utilize, utilise, apply, employ (put into service; make work or employ for a particular purpose or for its inherent or natural purpose)  
*"use your head!"; "we only use Spanish at home"; "I can't use this tool"; "Apply a magnetic field here"; "This thinking was applied to many projects"; "How do you utilize this tool?"; "I apply this rule to get good results"; "use the plastic bags to store the food"; "He doesn't know how to use a computer"*
  - derivationally related form
  - sentence frame

# WordNet

Существительные и глаголы организованы в иерархии, определённой отношением IS-A.

На самом верхнем уровне эти иерархии организованы в 25 деревьев для существительных и 15 для глаголов.

```
dog, domestic dog, Canis familiaris
└ canine, canid
  └ carnivore
    └ placental, placental mammal, eutherian, eutherian mammal
      └ mammal
        └ vertebrate, craniate
          └ chordate
            └ animal, animate being, beast, brute, creature, fauna
              └ ...
```

# WordNet

WordNet используется для задач разрешения лексической многозначности, информационного поиска, автоматической классификации текстов, автоматического реферирования текстов, машинного перевода и т.д.

Пакеты:

- <http://www.d.umn.edu/~tpederse/similarity.html> - WordNet::Similarity in Perl;
- <http://www.nltk.org/> - NLTK in Python

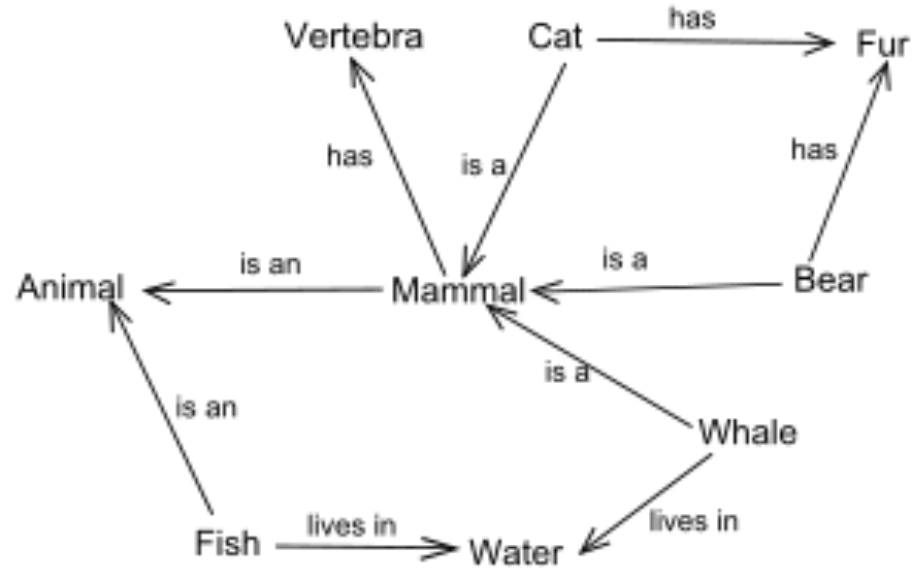
# Семантические сети

**Семантическая сеть** - это сеть, в которой представлены семантические отношения между концептами. Это направленный или ненаправленный граф, состоящий из узлов, которые являются концептами, и рёбер, которые являются семантическими отношениями между концептами.

Семантическая сеть используется в том случае, когда информацию можно представить как набор концептов и отношений между ними.

Семантические сети используются, например, для обнаружения плагиаризма.

# Semantic Network





# Семантические сети

Семантические отношения:

- синонимия;
- антонимия;
- A Kind Of, ISA;
- HasA;
- MemberOf, InstanceOf, Example;
- функциональные (*hasInfluenceOn*);
- количественные (*more, less*);
- пространственные (*far, near, under, on*);
- временные (*later, before*);
- логические (*AND, OR, NO*) etc.

# Семантические сети

1. У Пети есть отец по имени Иван.
2. Для Пети найдётся отец из множества мужчин.
3. Найдётся человек, для которого Иван - отец.
4. У каждого человека есть отец из множества мужчин.

1.  $\exists \text{ петя} \ \& \ \exists \text{ иван} : \text{отец}(\text{иван}, \text{петя});$
2.  $\exists \text{ петя} \rightarrow \exists x \in \text{мужчины} : \text{отец}(x, \text{петя});$
3.  $\exists \text{ иван} \rightarrow \exists y \in \text{люди} : \text{отец}(\text{иван}, y);$
4.  $\forall y \in \text{люди} \rightarrow \exists x \in \text{мужчины} : \text{отец}(x, y);$

# Семантические сети

## Недостатки:

- Семантическая сеть требует специального инструмента для формального вывода;
- Проблема нахождения решения задачи в семантической сети часто сводится к задаче нахождения сегмента в этой сети, который содержит необходимую информацию;
- Представление, использование и изменение информации - очень трудозатратный процесс

# Семантические сети

<http://www.cse.buffalo.edu/sneps/> - SNePS;

[http://pi7.fernuni-hagen.de/forschung/multinet/multinet\\_en.html](http://pi7.fernuni-hagen.de/forschung/multinet/multinet_en.html) - MultiNet

# Фреймы

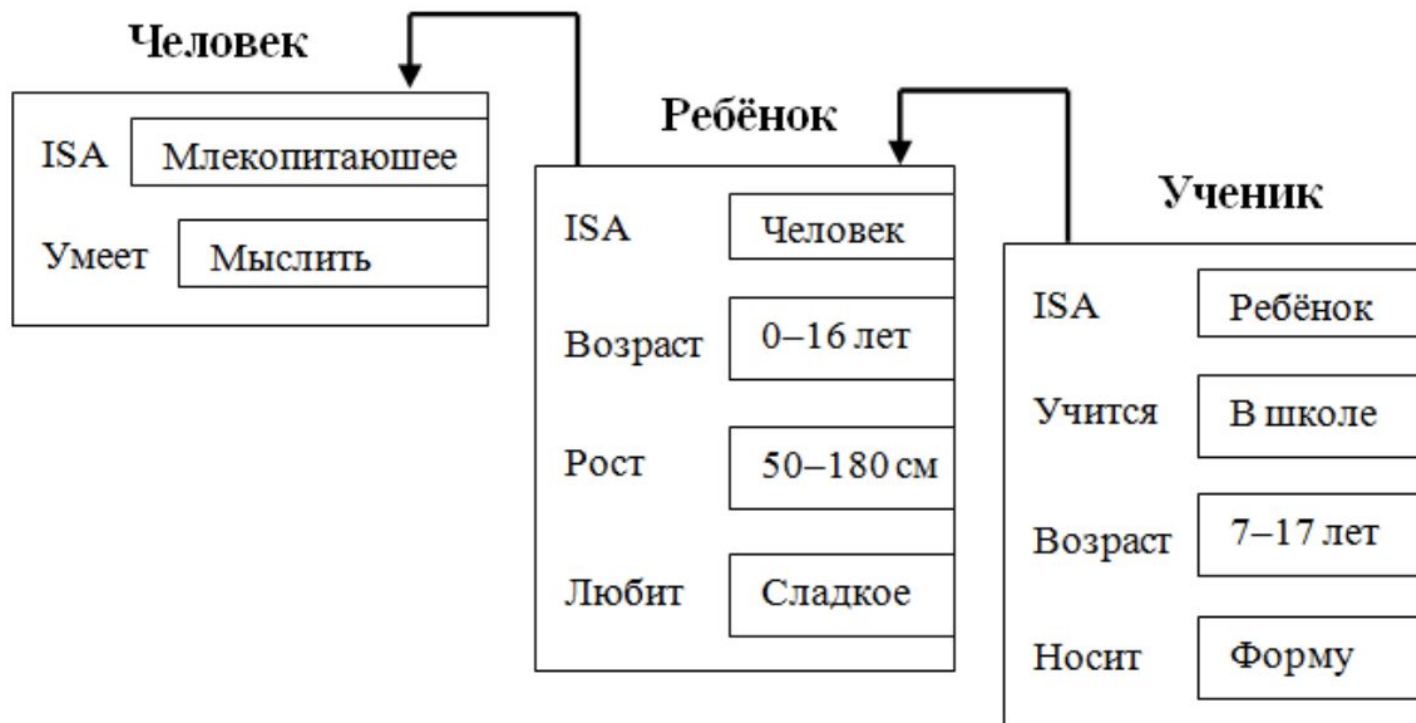
**Фрейм** (frame) — способ представления знаний в искусственном интеллекте, представляющий собой схему действий в реальной ситуации.

Фрейм — это модель абстрактного образа, минимально возможное описание сущности какого-либо объекта, явления, события, ситуации, процесса.

Особенность фреймовых моделей представления знаний состоит в том, что все понятия, описываемые в каждом из узлов модели, определяются набором атрибутов и их значениями, которые содержатся в слотах фрейма.

Фреймы используются в системах искусственного интеллекта (например, в экспертных системах) как одна из распространенных форм представления знаний.

# Фреймы



# Фреймы

Slot	Value	Type
ALEX	_	(This Frame)
NAME	Alex	(key value)
ISA	Boy	(parent frame)
SEX	Male	(inheritance value)
AGE	IF-NEEDED: Subtract(current,BIRTHDATE);	(procedural attachment)
HOME	100 Main St.	(instance value)
BIRTHDATE	8/4/2000	(instance value)
FAVORITE_FOOD	Spaghetti	(instance value)
CLIMBS	Trees	(instance value)
BODY_TYPE	Wiry	(instance value)
NUM_LEGS	1	(exception)

# Фреймы

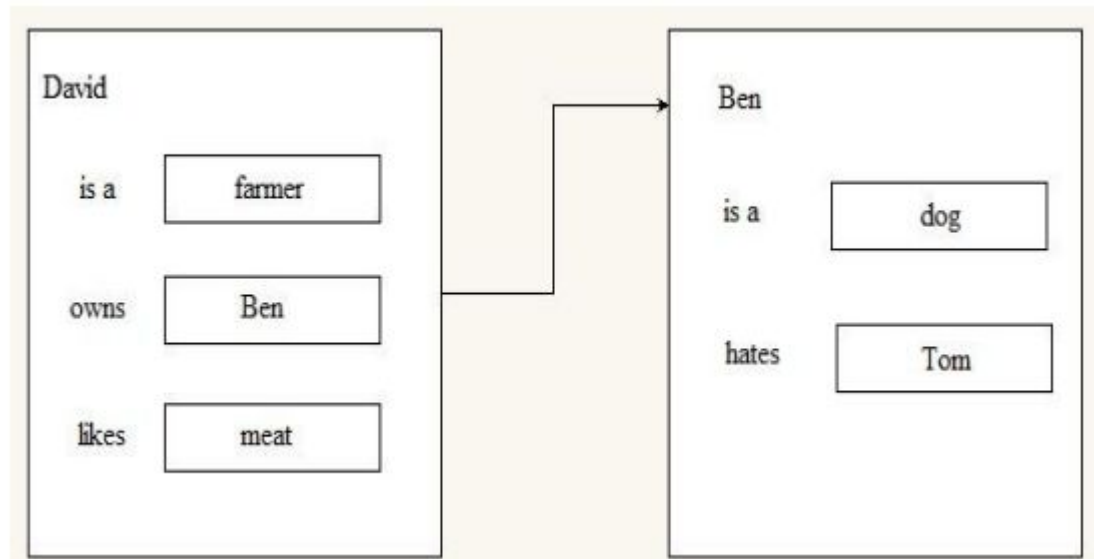
Slot	Value	Type
BOY	—	(This Frame)
ISA	Person	(parent frame)
SEX	Male	(instance value)
AGE	Under 12 yrs.	(procedural attachment - sets constraint)
HOME	A Place	(frame)
NUM_LEGS	Default = 2	(default, inherited from Person frame)



# Фреймы

Slot	Value	Type
MONKEY	–	(This Frame)
ISA	Primate	(parent frame)
SEX	OneOf(Male,Female)	(procedural attachment)
AGE	an integer	(procedural attachment - sets constraint)
HABITAT	Default = Jungle	(default)
FAVORITE_FOOD	Default = Bananas	(default)
CLIMBS	Trees	–
BODY_TYPE	Default = Wiry	(default)
NUM_LEGS	Default = 2	(default)

# Фреймы



# Фреймы

## Достоинства:

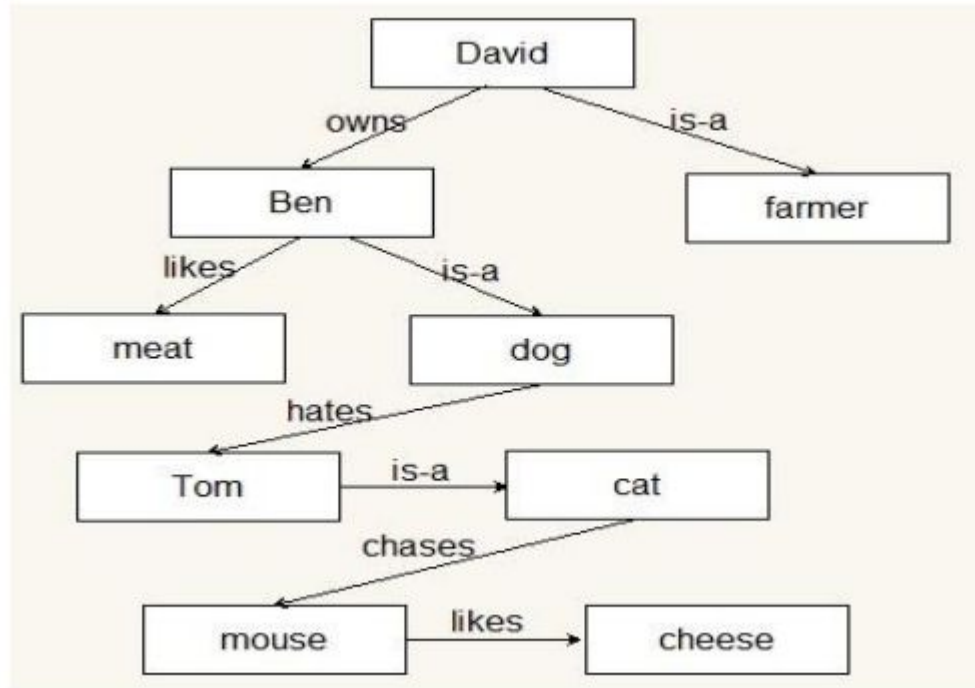
1. Фреймовое представление знаний облегчает программирование за счёт группировки связанных данных;
2. Фреймы интуитивно понятны во многих областях применения;
3. Фреймовое представление понятно и используется людьми, которые не являются ни программистами, ни дизайнерами системы;
4. Просто добавлять слоты для новых атрибутов и отношений;
5. Просто устанавливать дефолтные значения

# Фреймы

Недостатки:

- Отсутствие программного обеспечения

# Representation the Frame in the form of Semantic Network



*Semantic Network*

# Representation Semantic Network in the form of Frame

Frame name	Slot	Slot values
David	is-a	farmer
	Owns	Ben
	Likes	meat
Ben	is-a	dog
	Hates	Tom
Tom	is-a	cat
	Chases	mouse
Mouse	Likes	cheese

**Спасибо за внимание!**