Natural Language Processing

Lecture 5
Knowledge Representation for Computer Processing - Part 1

Knowledge Representation

Knowledge representation (KR) is the field of AI dedicated to representing information about the world in a form that a computer system can utilize to solve complex tasks.

KR is a key enabling technology for the Semantic web.

Knowledge representation:

- Logical models;
- Production models;
- Network models;
- Frame models;
- Ontology models.

Thesaurus

Thesaurus is a reference work that lists words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms), in contrast to a dictionary, which provides definitions for words, and generally lists them in alphabetical order.

Thesaurus

- **Synonyms** are words or phrases which mean exactly or nearly the same: *buy* and *purchase*, *big* and *large*, *quickly* and *speedily*, *on* and *upon*;
- **Antonyms** are words with opposite or nearly opposite meanings: *hot* and *cold*, *large* and *small*, *thick* and *thin*;
- **Hypernyms** and hyponyms are words that refer to, respectively, a general category and a specific instance of that category: *vehicle* is a hypernym of *car*, and *car* is a hyponym of *vehicle*;
- **Troponym**: a verb conveying a meaning that is a particular case of the meaning of another verb: *to duel* is a troponym of *to fight*; *to write* is a troponym of *to communicate*;
- **Meronym**: a word that names a part that belongs to and is therefore subordinate to a larger entity; a part-whole relationship, such as "*door*" or "*window*" in "*house*", or "*engine*" or "*steering-wheel*" in "*car*";

Thesaurus

- **Holonym** is a word for the whole of which other words are part: *house* contains *roof* and *window*;
- **Paronym** a word that is related to another word and derives from the same root: *dubious* and *doubtful*.

Thesaurus: WordNet

http://wordnet.princeton.edu/

WordNet is a lexical database for the English language. It groups English words into sets of synonyms called **synsets**, provides short definitions and usage examples, and records a number of relations among these synonym sets of their members.

- WordNet interlinks not just word forms - strings or letters - but specific senses of words → words that are found in close proximity to one another in the network are semantically disambiguated;
- WordNet labels the semantic relations among words.

WordNet Search - 3.1

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

Word to search for: wordnet [Search WordNet]

Display Options: (Select option to change) ▾ [Change]

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)

Thesaurus: WordNet

The database contains 155,827 words organized in 117,659 synsets for a total of 206,941 word-sense pairs.

WordNet includes nouns, verbs, adjectives and adverbs.

Words from the same lexical category that are roughly synonymous are grouped into synsets. Synsets include both words and collocations (e.g. '*eat out*'). The different senses of a polysemous word form are assigned to different synsets. The meaning of a synset is further clarified with a short defining gloss and one or more usage examples.

Thesaurus: 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*

Thesaurus: WordNet

Both nouns and verbs are organized into hierarchies, defined by hypernym or **IS A** relationships.

At the top level, these hierarchies are organized into 25 beginner "trees" for nouns and 15 for verbs, which are called **lexicographic files**.

```
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
                              └ ...
```

Thesaurus: WordNet

WordNet has been used for word-sense disambiguation, information retrieval, automatic text classification, automatic text summarization, machine translation, automatic crossword puzzle generation etc.

Packages:

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

Semantic Network

A semantic network is a network that represents semantic relations between concepts. It is a directed or undirected graph consisting of vertices, which represent concepts, and edges, which represent semantic relations between concepts.

A semantic network is used when one has knowledge that is best understood as a set of concepts that are related to one another.

Semantic networks are used in specialized information retrieval tasks, such as plagiarism detection.

Semantic Network
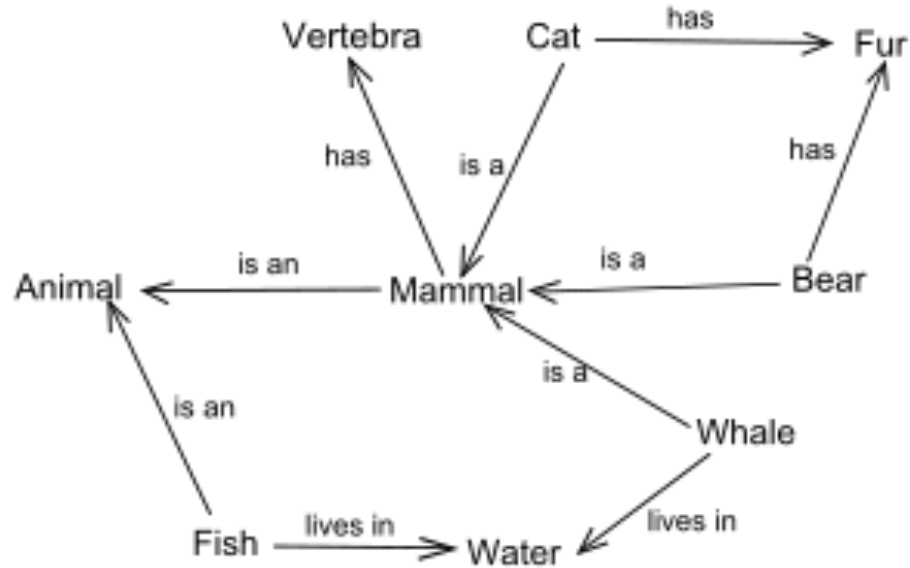
Semantic Network

Semantic relations:

- synonymy;
- antonymy;
- A Kind Of, ISA;
- HasA;
- MemberOf, InstanceOf, Example;
- functional (*hasInfluenceOn*);
- quantitative (*more*, *less*);
- dimensional (*far*, *near*, *under*, *on*);
- temporal (*later*, *before*);
- logical (*AND*, *OR*, *NO*) etc.

Semantic Network

1. John has a father named Peter.
2. John has a father from a set of men.
3. There is a person, who has a father John.
4. Every person has a father from a set of men.

1. ∃ John & ∃ Peter : father(Peter, John);
2. ∃ John → ∃ x ∈ men : father (x, John);
3. ∃ John → ∃ y ∈ humans : father(John, y);
4. ∀ y ∈ humans → ∃ x ∈ men : father (x, y);

Semantic Network

Disadvantages:

- Semantic network doesn't contain clear representation about knowledge domain structure, so creation and modification are difficult;
- Semantic network requires a special instrument for formal deduction;
- The problem of finding solution in semantic network is a task to find a segment in this network which contains required information;
- Representation, using and modification of knowledge is a time-consuming process.

Semantic Network

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

http://pi7.fernuni-hagen.de/forschung/multinet/multinet_en.html - MultiNet

Frame

A frame is an artificial intelligence data structure used to divide knowledge into substructures by representing "stereotyped situations".

The frame contains information on how to use the frame, what to expect next, and what to do when these expectations are not met.

Some information in the frame is generally unchanged while other information, stored in "terminals", usually change.

Frame

Each piece of information about a particular frame is held in slot. The information can contain:

- Facts or Data;
  - Values (called facets);
- Procedures;
  - IF-NEEDED: deferred evaluation;
  - IF-ADDED: updates linked information;
- Default values:
  - For Data;
  - For Procedures;
- Other Frames or Subframes

Frame

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)

Frame

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)

Frame

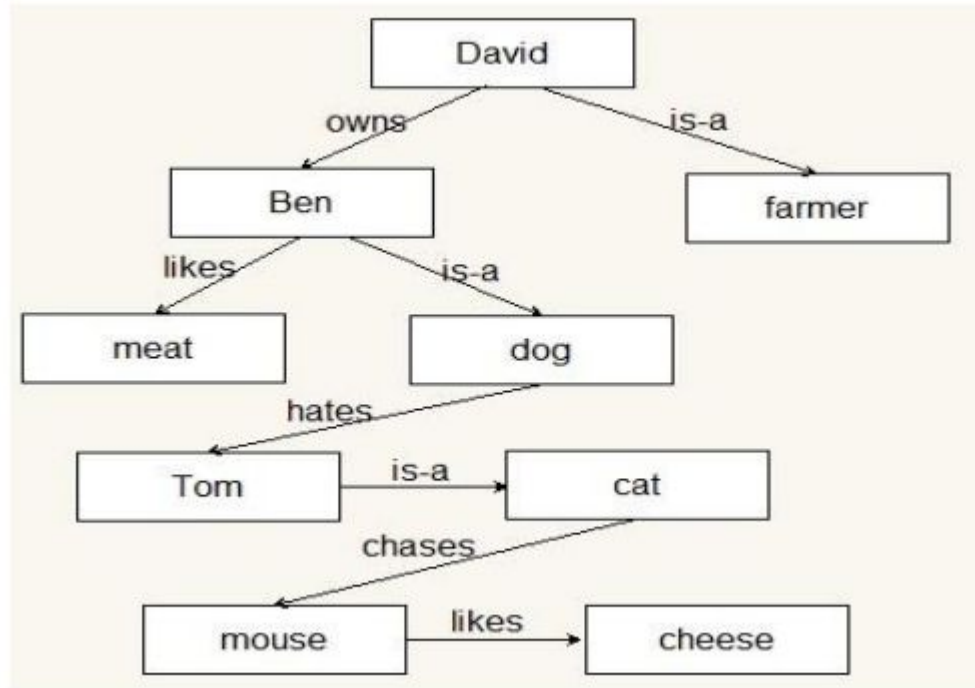| Slot	Value	Type
MONKEY	_	(This Frame)
ISA	Primate	(parent frame)
SEX	OneOf(Male,Female)	(procedural attachment)
AGE	an integer	(procedural attachment - sets constraint)
HABITAT	Default = Jungle	(default)
FAVORITE_FOOD	Default = Bananas	(default)
CLIMBS	Trees	_
BODY_TYPE	Default = Wiry	(default)
NUM_LEGS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

Frame
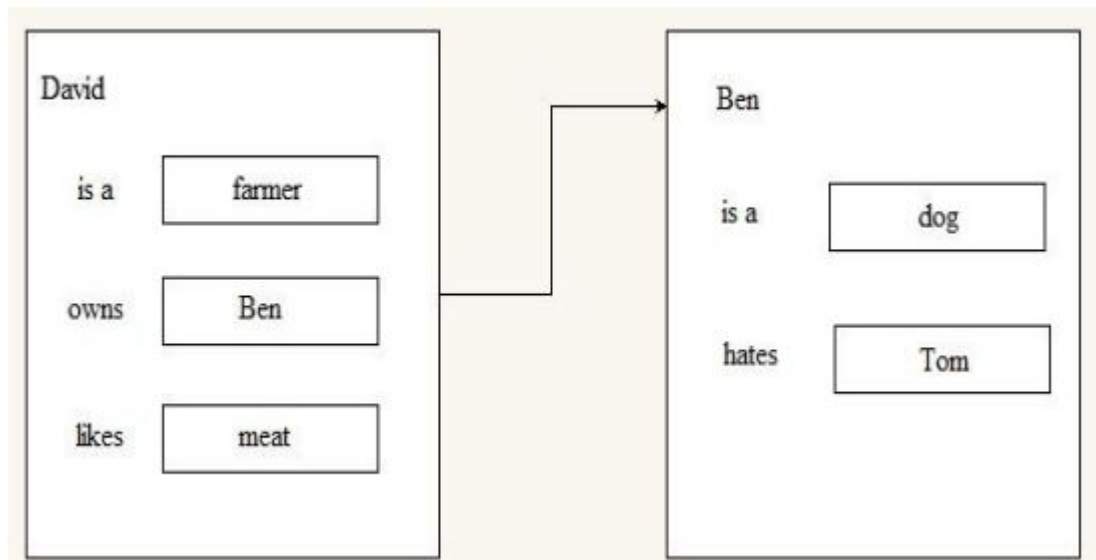
Frame

Advantages:

1. The frame knowledge representation makes the programming simpler by grouping related data;
2. The frame is flexible and intuitive in many application areas;
3. The frame representation is easily understood and used by people who are neither programmer nor designer of a system;
4. It is not hard to add slots for new attributes and relations;
5. It is simple to include default data and to discover the missing values.

Frame

Disadvantages:

- It is difficult to use the frame system in a program, so the algorithm is required in the process of using the frame in the program;
- The lack of low-priced computer software;
- Inference mechanism is not easily processed in a frame system.

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

Thank you for your attention!