KDI ⋮ Knowledge and Data Integration

Wordnet

Phase: 1.   Language & Standards

W8.L15.T15.1.3

Contents

Contents

Motivation

Let us take a look at the following natural language processing problem,

Problem (Information retrieval)

Task of retrieving the subset of documents in the large electronic collection based on the query by user which may be anything from a single word to a paragraph.

However, we can run into issues when any strategy that simply uses string matching to select documents,

1. Polysemy, many words have multiple meanings which contain terms also found in the user's query is bound to return many irrelevant documents.
2. Synonymy, many equivalent or closely related meanings can be conveyed by distinct words which also can be found in the user's query is bound to miss many relevant documents.

Solution : Wordnet

A formal solution known as the wordnet was developed,

- a **network** containing group of similar words, i.e.synsets which are provided with a brief description known as gloss,
- the **node** of the network, i.e.concept which is expressed by a group of synonyms(synsets) i.e.sense based on part-of-speech,
- and the **lines** connecting these nodes(concepts) are semantic relations such as hyponymy, antonymy, part-whole, cause-effect etc.
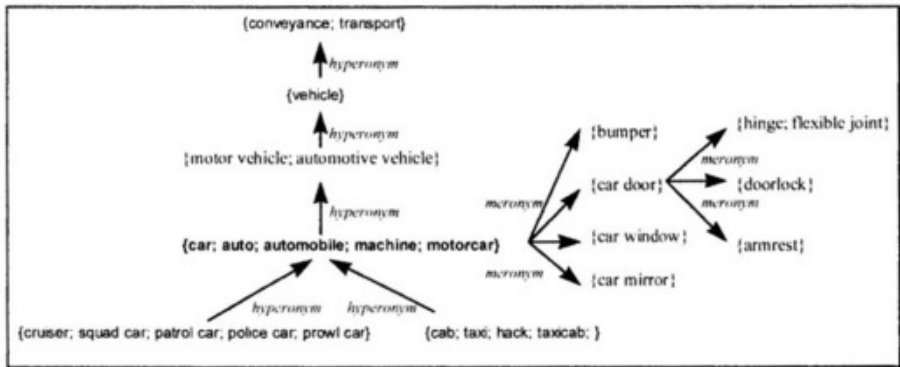
Example

Synset referring to the same concept : {car; auto; automobile; machine; motorcar}

synset described by a **gloss**: "4-wheeled; usually propelled by an internal combustion engine".

semantic relation: {cab; taxi; taxicab} is a hyponym

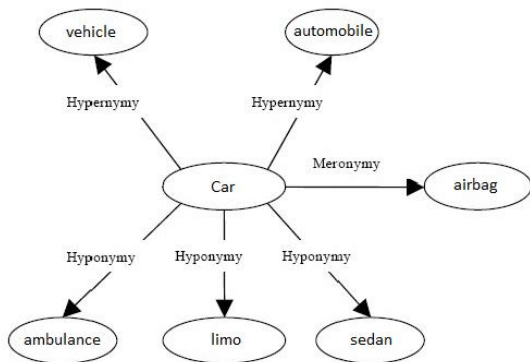# Representation of wordnet as synsets

Contents

Terms and Terminology

The other important lexical relations based on part-of-speech used in representation of wordnet are,

- noun :
    - synonym - antonym
    - hypernyms(ex. *color* is the hyperenym of *red*) - hyponym (ex. *bowl* is a hyponym of *dish*)
    - holonym(ex. *building* is a holonym of *window*) - meronym(ex. *beak* and *wings* are meronym of *bird*)
- verb :
    - synonym - antonym
    - hypernym - troponym (ex. *to lisp* is a troponym *to talk*)
    - entailment (ex. *to snore* entails *to sleep*)
- adjective : modifying the nouns, ex. *heavy* is the opposite of *light*
- adverb : derived from adjectives with suffixation, for ex. *quickly* is the opposite of *slowly*

Representation of wordnet as a network

This is an example of noun sense which is represneted as a graph with nodes(concepts) and lines connecting (relations),

Example of wordnet UI

WordNet Search - 3.1
- **WordNet home page** · **Glossary** · **Help**

Word to search for: [cycle] [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) **cycle**, rhythm, round (an interval during which a recurring sequence of events occurs) *"the never-ending cycle of the seasons"*
- **S:** (n) **cycle** (a series of poems or songs on the same theme) *"Schubert's song cycles"*
- **S:** (n) **cycle** (a periodically repeated sequence of events) *"a cycle of reprisal and retaliation"*
- **S:** (n) hertz, Hz, cycle per second, cycles/second, cps, **cycle** (the unit of frequency; one hertz has a periodic interval of one second (named for Heinrich Rudolph Hertz))
- **S:** (n) **cycle**, oscillation (a single complete execution of a periodically repeated phenomenon) *"a year constitutes a cycle of the seasons"*
- **S:** (n) bicycle, bike, wheel, **cycle** (a wheeled vehicle that has two wheels and is moved by foot pedals)

Verb

- **S:** (v) **cycle** (cause to go through a recurring sequence) *"cycle the laundry in this washing program"*
- **S:** (v) **cycle** (pass through a cycle) *"This machine automatically cycles"*
- **S:** (v) motorbike, motorcycle, **cycle** (ride a motorcycle)
- **S:** (v) bicycle, **cycle**, bike, pedal, wheel (ride a bicycle)
- **S:** (v) **cycle** (recur in repeating sequences)

Size of wordnet

The wordnet 3.1 is the latest version of the Princeton wordnet.

	Princeton WordNet 3.1
Synsets	117,791
Lemma	159,015
Senses	207,272
Synset Relations	285,668
Sense Relations	92,535
Definitions	117,791
Examples	47,539

POS	Unique Strings	Synsets	Total Word-Sense Pairs
Noun	117798	82115	146312
Verb	11529	13767	25047
Adjective	21479	18156	30002
Adverb	4481	3621	5580
Totals	155287	117659	206941

Contents

Uses of wordnet

Recall the defintion of the wordnet,

Definition (Wordnet)

WordNetR is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept.

With the above definition, we can list the uses based on each feature of this large resource,

- as an online dictionary because it is a lexical resource
- for the task of information retrieval i.e. selection of textual information pertinent to users' queries;

- for the task of word sense disambiguation because of organisation of data in wordnet (set of words with similar meaning)

  Ex : use of word bank in the following sentences
  1. Deposit the money in the bank
  2. A bank of clouds was building to the northeast

- for automatic text classification, where documents are assigned to one or more thematic categories depending on their content; in this the conceptual notion of wordnet is the key.

- for machine translation as it can be extended to multiple languages. It is even utilized by Google Translate as a part of the translation process between languages which possess WordNets.

Contents

Monolingual to Multilingual

Wordnet is a monolingual lexical resource, however it led to development of lexical resources in various languages

- most widespread and best known ones such as French, German, Chinese, Spanish or Portuguese
- minority languages too, such as Albanian, Kannada, Croatian, Basque, Catalan and Galician.

Some other multilingual projects are,

- EuroWordnet : consisting of european languages wordnets such as Dutch, Spanish and Italian, which are linked to the English wordnet.
- MultiWordNet : multilingual lexical database in which the Italian WordNet is strictly aligned with Princeton WordNet 1.6.

References

- Princeton wordnet : https://wordnet.princeton.edu/
- MultiWordNet : https://multiwordnet.fbk.eu/online/multiwordnet.php

W8.L15.T15.1.3

Wordnet

Phase: 1.   Language & Standards