

NLP - MODULE 4 - CHAPTER 4

Semantic Analysis

Semantic Analysis

- process of extracting meaning from text.

- imp subtask of NLP
- divided in 2 parts
 - 1) Studying meaning of individual word
 - 2) Studying combination of individual words

Lexical Semantics

- relationship between lexical items,

meaning of sentences, and syntax of sentence

- includes 2 points:

- classification & decomposition of lexical items
- difference & similarities between numerous lexical semantic structures

Elements of Lexical Semantics

Analysis

- 1) **Hyponymy**:
 - specific lexical items of a general lexical item (hyperlink)
 - eg: apple is hyponym of fruit
- 2) **Meronymy**:
 - logical arrangement of text & word that represent a part of / member of something
 - eg: segment of an apple

- 3) **Polysemy**:
 - relationship between meaning of word or phrases, although slightly different, they share a common basic meaning
- 4) **Synonymy**:
 - words that have same sense or meaning
 - eg: Sad, unhappy, depressed
- 5) **Antonymy**:
 - words that have opposite meanings
 - eg: good / bad
- 6) **Homonymy**:
 - words that sound same & spelled same but have different meaning
 - eg: Right (Correct) / Right (Turn)

- 1) **Hyponymy**:
 - specific lexical items of a general lexical item (hyperlink)
 - eg: apple is hyponym of fruit
- 2) **Meronymy**:
 - logical arrangement of text & word that represent a part of / member of something
 - eg: segment of an apple

- 3) **Polysemy**:
 - relationship between meaning of word or phrases, although slightly different, they share a common basic meaning
- 4) **Synonymy**:
 - words that have same sense or meaning
 - eg: Sad, unhappy, depressed
- 5) **Antonymy**:
 - words that have opposite meanings
 - eg: good / bad
- 6) **Homonymy**:
 - words that sound same & spelled same but have different meaning
 - eg: Right (Correct) / Right (Turn)

- 1) **Hyponymy**:
 - specific lexical items of a general lexical item (hyperlink)
 - eg: apple is hyponym of fruit
- 2) **Meronymy**:
 - logical arrangement of text & word that represent a part of / member of something
 - eg: segment of an apple

- 3) **Polysemy**:
 - relationship between meaning of word or phrases, although slightly different, they share a common basic meaning
- 4) **Synonymy**:
 - words that have same sense or meaning
 - eg: Sad, unhappy, depressed
- 5) **Antonymy**:
 - words that have opposite meanings
 - eg: good / bad
- 6) **Homonymy**:
 - words that sound same & spelled same but have different meaning
 - eg: Right (Correct) / Right (Turn)

- 1) **Hyponymy**:
 - specific lexical items of a general lexical item (hyperlink)
 - eg: apple is hyponym of fruit
- 2) **Meronymy**:
 - logical arrangement of text & word that represent a part of / member of something
 - eg: segment of an apple

- 3) **Polysemy**:
 - relationship between meaning of word or phrases, although slightly different, they share a common basic meaning
- 4) **Synonymy**:
 - words that have same sense or meaning
 - eg: Sad, unhappy, depressed
- 5) **Antonymy**:
 - words that have opposite meanings
 - eg: good / bad
- 6) **Homonymy**:
 - words that sound same & spelled same but have different meaning
 - eg: Right (Correct) / Right (Turn)

Semantic Attachment

- process of making semantics of a sentence by attaching pieces of semantics to the syntax tree

- 1) Semantic specialist
 - Their job is to create complex semantic structure that represent the sentence's
- 2) Lambda Calculus
 - It extend first order predicate calculus with function application
 - The predicate calculus (PC) includes unary, binary & n-ary predicated
 - 3) Feature Unification
 - it's a partial operation of feature structure

- it's the operation of combining two feature structure such that the new feature structure contains all the info of the original & nothing more

Strategy for Semantic Attachment

- 1) Create complex, lambda expressions with lexical items
- 2) Introduce quantifiers, predicates, terms

- 3) Percolate up semantics from child if non-branching
- 4) Apply semantics of one child to other through lambda
- 5) Combine elements, but don't introduce

Senses

- One word can have multiple meaning
- The unit of meaning is sense

- They are lexemes that share a form but unrelated meaning
- words that have same syntax / spelling / form but meaning are different
- eg: bank (financial institution) / bank (riverbank)

Polysemy

- words or phrases with different but related meaning
- eg: he drank a glass of milk. / he forget to milk the cow

Synonymy

- word or phrase that means exactly the same as another word / phrase in same language

- words with similar meaning
- eg: rich-wealthy, mistake-error

Antonymy

- words that have opposite or contrasting meaning
- eg: rich/poor, hot/cold, day/night

Hypernymy & Hyponymy

- **Hyponymy** is the sense which is subclass of another sense
- eg: car is hyponym of vehicle
- dog — animal
- mango — fruit

- **Hypernymy** is the sense which is superclass
- eg: vehicle is the hypernym of car
- animal — dog
- fruit — mango

- Hyponyms & Hypernyms are Asymmetric

Meronymy

- It is a word that represents a constituent part or member of something
- eg: mango is meronym of mango tree
- mango < mango tree

Word Net & Synset

- Word net is a bit collection of word from english language that are related to each other and are grouped in some way
- also called lexical database

- Wordnet is a database of english language that are connected together by their semantic relation

- It is like super set dictionary with graph structure

- Word net groups nouns, verbs, adjectives which are similar and groups are called synset or synonym

- eg: Synset {Stone, cement} & synset of building materials

- 3 principle for Synset
 - 1) minimality
 - 2) coverage
 - 3) replaceability

Word Sense Disambiguation (WSD)

- Well known problem in NLP
- used in identifying what sense of word meant in a sentence when word has multiple meanings.

- Approaches to WSD:
 - 1) Dictionary & knowledge based method
 - 2) supervised method
 - 3) Semi supervised / minimally supervised method
 - 4) Unsupervised method

WSD Evaluation

- require test corpus, hand annotated with target or correct senses & assume that such a corpus can be constructed.

- performance measures
 - 1) Precision
 - 2) Recall
- Types of test corpus
 - 1) lexical sample
 - 2) All words

Difficulties in WSD

- 1) Difference between dictionaries
- 2) Interchange Variance
- 3) Different Algorithm for different app
- 4) Word sense distinctions

Applications of WSD

- 1) Machine Translation
- 2) Text mining = Information Extraction (eg)
- 3) Information Retrieval (IR)
- 4) Lexicography

Dictionary (Knowledge Based Approach)

- Ways to implement this approach:
 - maximal matching

- bidirectional maximal matching
- maximum matching

Difference between Polysemy & Homonymy

Polysemy

- 1) 2 words or phrases have many possible meanings

- 2) Different but related meaning
- 3) Related word origin
- 4) words are entered under one entry in dictionary
- 5) can be understood if one knows the meaning of word

Homonymy

- 1) Two unrelated words that look same / sound same have different meaning.

- 2) Different meanings.
- 3) Different word origin
- 4) words are entered separately in dictionary
- 5) cannot be understood as words have unrelated different meanings

