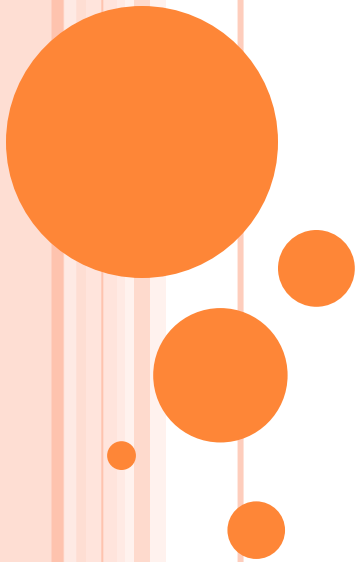


SEMANTIC ANALYSIS MODULE 4



INTRODUCTION

- Semantic analysis is the process of finding the meaning from text.
- This analysis gives the power to computers to understand and interpret sentences, paragraphs, or whole documents, by analyzing their grammatical structure, and identifying the relationships between individual words of the sentence in a particular context.
- Therefore, the goal of semantic analysis is to draw exact meaning or dictionary meaning from the text. The work of a semantic analyzer is to check the text for meaningfulness.



- **How is Semantic Analysis different from Lexical Analysis?**
- Lexical analysis is based on smaller tokens whereas, the semantic analysis focuses on larger chunks.
- Since semantic analysis focuses on larger chunks, therefore we can divide the semantic analysis into the following two parts:
- **Studying the meaning of the Individual Word**
- It is the first component of semantic analysis in which we study the meaning of individual words. This component is known as lexical semantics.
- **Studying the combination of Individual Words**
- In this component, we combined the individual words to provide meaning in sentences. This is known as Compositional Semantics.
- Eg: I like you.
- You like me.



↓ Lexical Semantics

Study of meanings of individual words.

words have meanings, internal structure, & are involved in diff. relationship with other words.

Concerned with the linguistic study of systematic meaning, the way in which related structure of words or lexemes.

Eg:-

I like you & You like me. — Same words but different meanings in the sentence. This is because in the 1st sent 'I' = subject 'You' = object but in 2nd sent these are reversed.

o This means that the overall meaning relies not only on the meaning of each part, but also on syntactic composition.

o This is also known as principle of compositionality.

↓ Compositional Semantics

• Study of combination of individual words.

o I don't eat I don't drink.
; we can make diff blw these sentences because of the lexical meaning of 'eat' & 'drink'.

However, syntax or the way in which the sentence is constructed is also important.

- **Example**, consider the following sentence:
- **Sentence: Ram is great**
- In the above sentence, the speaker is talking either about Lord Ram or about a person whose name is Ram. That is why the task to get the proper meaning of the sentence is important.
- Example:
- “Your customer service is a joke! I have been on hold for 30 minutes and counting!”
- Any human can understand a customer is frustrated because a customer service agent is taking too long to respond.
- But machine cant understand the context in which the words are used.
- It might misinterpret the word “joke” as positive



MEANING REPRESENTATION

- Semantic analysis creates a representation of the meaning of a sentence.
- Building Blocks of Semantic System
- In word representation or representation of the meaning of the words, the following building blocks play an important role –
- **Entities** – It represents the individual such as a particular person, location etc. For example, Haryana. India, Ram all are entities.
- **Concepts** – It represents the general category of the individuals such as a person, city, etc.
- **Relations** – It represents the relationship between entities and concept. For example, Ram is a person.
- **Predicates** – It represents the verb structures. A predicate is the part of a sentence, or a clause, that tells what the subject is doing or what the subject is.
- Eg: The exam was difficult.
 Subject Predicate



- Approaches to Meaning Representations
- Semantic analysis uses the following approaches for the representation of meaning –
- First order predicate logic (FOPL)
- Semantic Nets
- Frames
- Conceptual dependency (CD)
- Rule-based architecture



- Predicate logic / FOPL: Also known as quantificational logic, and first-order predicate calculus.
- Predicates are properties, additional information to better express the subject of the sentence.

- ▶ It is an extension to propositional logic.

- ▶ It is represented with **quantifiers**.

- ▶ Example:

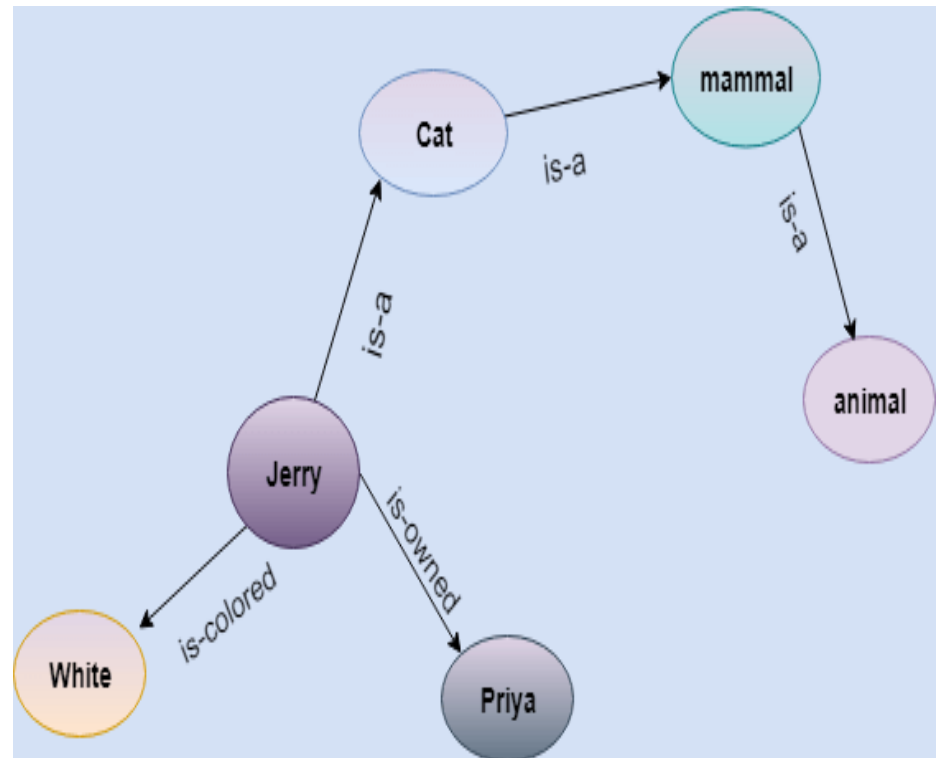
- ▶ John and Richard are brothers \rightarrow **Brothers(John, Richard)**

- ▶ All humans likes Ice-cream $\rightarrow \forall x \text{ humans}(x) \rightarrow \text{likes}(x, \text{Ice-cream})$

- Semantic Networks
- Semantic networks are an alternative to predicate logic as a form of knowledge representation. The idea is that we can store our knowledge in the form of a graph, with nodes representing objects, and arcs representing relationships between those objects.

- **Statements:**

1. Jerry is a cat.
2. Jerry is a mammal
3. Jerry is owned by Priya.
4. Jerry is white colored.
5. All Mammals are animal.



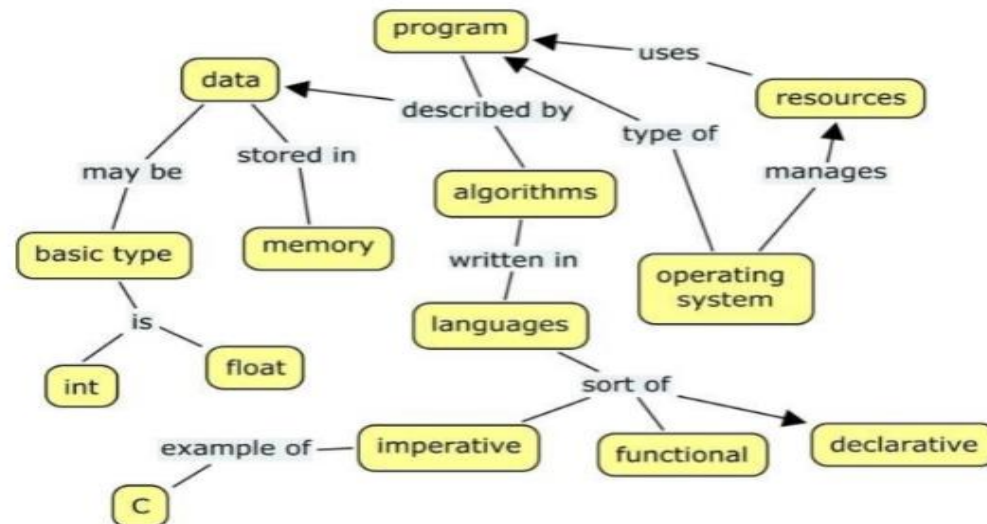
- Frames:
- A frame is a record like structure which consists of a collection of attributes and its values to describe an entity in the world.
- It consists of a collection of slots and slot values (Fillers).
- Example: The car is the object, the slot name is the attribute, and the filler is the value.

Slots	Fillers
manufacturer	General Motors
model	Chevrolet Caprice
year	1979
transmission	automatic
engine	gasoline
tires	4
color	blue



- Conceptual Dependency Diagrams / Concept Maps
- A concept map (CM) is a graphical tool that has been successfully used for organizing and representing knowledge. It includes concepts, usually signified by nouns or noun phrases, and relationships between them indicated by a line linking two concepts.
- Labeling a line with a verb or a verb phrase creates a concept-label-concept chain that can be read as a sentence.

Example



Concept Map extracted from 'Operating System'

- Rule-based Systems
- Rule-based systems provide the computational mechanisms found in most expert systems. Knowledge is specified via facts and IF–THEN rules.
- IF(Condition) THEN (Action)
- Example:
 - **IF (at bus stop AND bus arrives) THEN action (get into the bus)**
 - **IF (on the bus AND paid AND empty seat) THEN action (sit down).**
 - **IF (on bus AND unpaid) THEN action (pay charges).**
 - **IF (bus arrives at destination) THEN action (get down from the bus).**



LEXICAL SEMANTICS:

- The first part of semantic analysis, studying the meaning of individual words is called lexical semantics.
- It includes words, sub-words, affixes (sub-units), compound words and phrases also. All the words, sub-words, etc. are collectively called lexical items.
- In other words, we can say that lexical semantics is the relationship between lexical items, meaning of sentences and syntax of sentence.
- Following are the steps involved in lexical semantics –
- Classification and Decomposition of lexical items like words, sub-words, affixes, etc. is performed in lexical semantics.
- Differences as well as similarities between various lexical semantic structures is also analyzed.



○ **Elements of Semantic Analysis/Relations Among Lexemes and their senses.**

- Followings are some important elements of semantic analysis –

○ **Hyponymy**


- It may be defined as the relationship between a generic term and instances of that generic term.
- They can be understood by taking class-object as an analogy.
- Here the generic term is called hypernym and its instances are called hyponyms.
- For example, the word color is hypernym and the color blue, yellow etc. are hyponyms.



○ **Homonymy**

- Homonymy refers to two or more lexical terms with the same spellings but completely distinct in meaning.
- For example, the word “Bat” is a homonymy word because bat can be an implement to hit a ball or bat is a flying mammal also.
- For example: ‘Rose’ might mean ‘the past form of rise’ or ‘a flower’, – same spelling but different meanings; hence, ‘rose’ is a homonymy.

○ **Polysemy**

- Polysemy refers to lexical terms that have the same spelling but multiple closely related meanings.
 - It differs from homonymy because the meanings of the terms need not be closely related in the case of homonymy.
- 

- For example: ‘man’ may mean ‘the human species’ or ‘a male human’ or ‘an adult male human’ – since all these different meanings bear a close association, the lexical term ‘man’ is a polysemy.
- For example: the word “bank” is a polysemy word having the following meanings –
 - A financial institution.
 - The building in which such an institution is located.
- **Synonymy**
 - When two or more lexical terms that might be spelt distinctly have the same or similar meaning, they are called Synonymy.
 - For example: (Job, Occupation), (Large, Big), (Stop, Halt).



○ **Antonymy**

- Antonymy refers to a pair of lexical terms that have contrasting meanings – they are symmetric to a semantic axis.
- For example: (Day, Night), (Hot, Cold), (Large, Small).

○ **Meronymy**

- Meronymy refers to a relationship wherein one lexical term is a constituent of some larger entity. For example: 'Wheel' is a meronym of 'Automobile'
- Meronymy is also a logical arrangement of text and words that denotes a constituent part of or member of something under elements of semantic analysis.
- Example: hand -- body.



CORPUS STUDY

- A corpus is a large and structured set of machine-readable texts that have been produced in a natural communicative setting.
- They can be derived in different ways like text that was originally electronic, transcripts of spoken language and optical character recognition, etc.
- Language is infinite but a corpus has to be finite in size.
- For the corpus to be finite in size, we need to sample and proportionally include a wide range of text types to ensure a good corpus design.



- The size of the corpus depends upon the purpose for which it is intended as well as on some practical considerations as follows –
- For Example: Medical Corpus for Medical Study
- Movies Corpus for Critics and rating.



WORDNET

- WordNet is a large lexical database of words, senses, and their semantic relations.
- WordNet captures the word and their senses.
- This means WordNet represents words (or senses) as lists of the word senses that can be used to express the concept.

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

- A hierarchically organized lexical database
- A machine-readable thesaurus, and aspects of a dictionary
- Versions for other languages are under development

part of speech	no. synsets
noun	82,115
verb	13,767
adjective	18,156
adverb	3,621


- In WordNet, the sense is defined by a set of synonyms, called synsets, that have a similar meaning or sense.

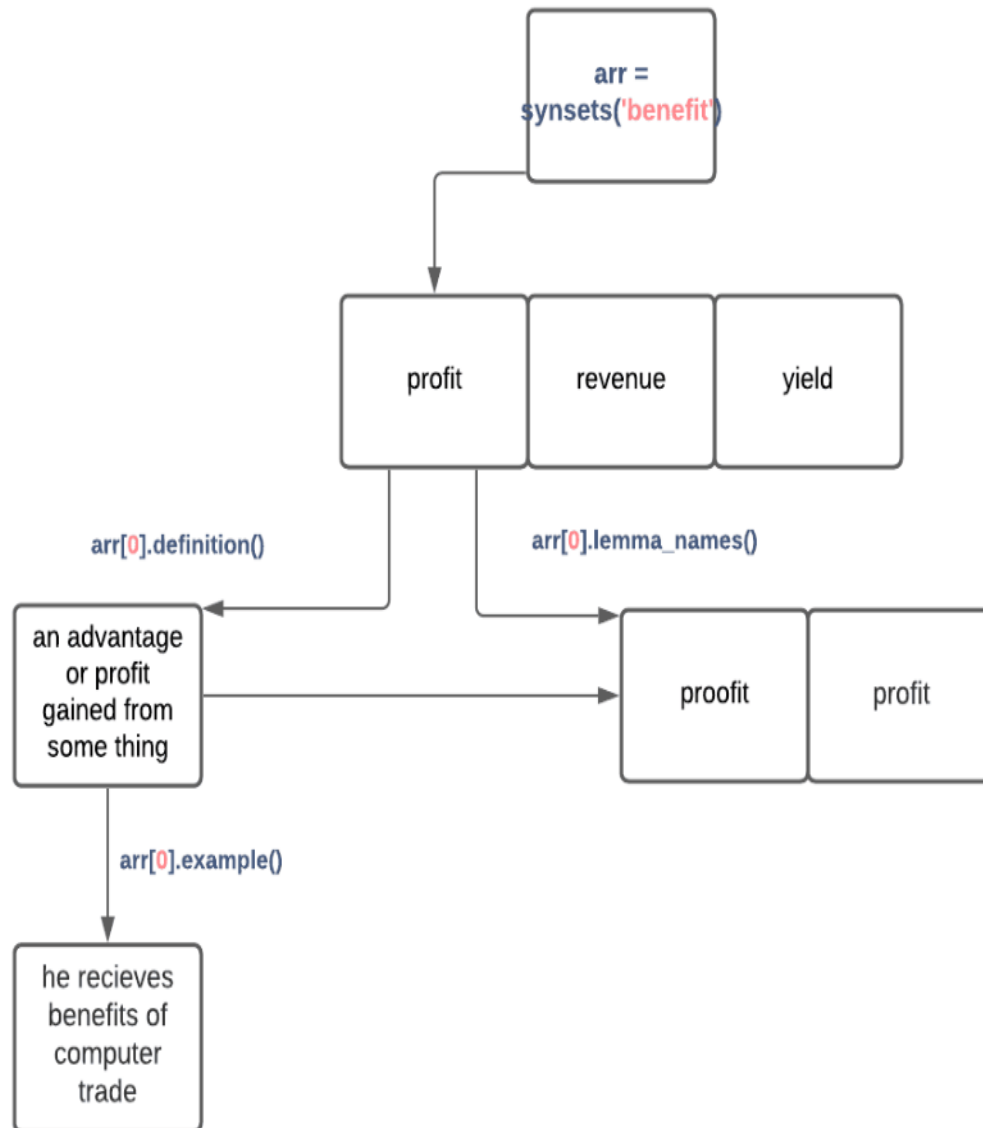
- A **synset** is a set of synonyms representing a sense
- Example: chump as a noun to mean 'a person who is gullible and easy to take advantage of'

{chump¹, fool², gull¹, mark⁹, patsy¹, fall guy¹, sucker¹,
soft touch¹, mug²}

- Each of these senses share this same gloss.
- For WordNet, the meaning of this sense of chump is this list.

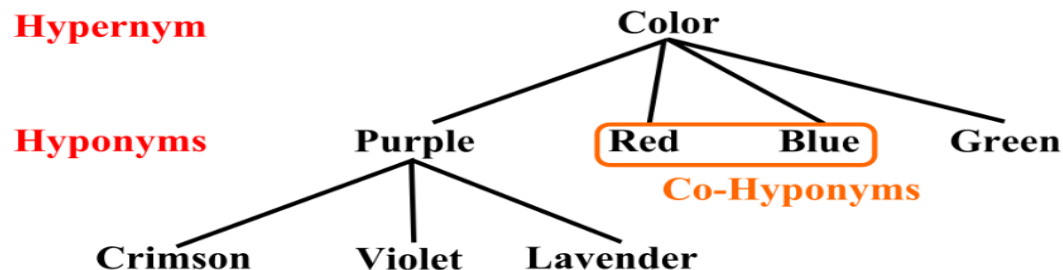
STRUCTURE OF WORDNET

- The main concept of the relationship between the words in the WordNETs network is that the words are synonyms like sad and unhappy, benefit and profit.
 - These words show the same concept of using them in similar contexts by interchanging them.
 - These types of words are grouped into synsets which are unordered sets. Where synsets are linked together if they are having even small conceptual relations.
 - Every synset in the network has its own brief definition and many of them are illustrated with the example of how to use them in a sentence.
 - That definition and example part makes WordNET different from other
- 



- **Relations in the WordNET**

- **Hyponym:**



- **Meronymy:**

- Example a bike has two wheels handle and petrol tank.


- **Troponymy:**

- Troponymy is the presence of a 'manner' relation between two lexemes. In WordNET Verbs describing events that necessarily and unidirectionally entail one another are linked: {buy}-{pay}, {succeed}-{try}, {show}-{see}, etc.

- **Antonymy:**



BABELNET

- BabelNet is a multilingual lexicalized semantic network and ontology.
 - BabelNet was automatically created by linking Wikipedia to the most popular computational lexicon of the English language, WordNet.
 - BabelNet (version 5.0) covers 500 languages.
 - It contains almost 20 million synsets and around 1.4 billion word senses.
 - Each Babel synset contains 2 synonyms per language, i.e., word senses, on average.
 - The semantic network includes all the lexico-semantic relations from WordNet (hypernymy and hyponymy, meronymy and holonymy, antonymy and synonymy, etc., totaling around 364,000 relation edges) as well as an underspecified relatedness relation from Wikipedia (totaling around 1.3 billion edges).
- 

Ambiguity and Uncertainty in Language

- Ambiguity, generally used in natural language processing, can be referred as the ability of being understood in more than one way.
- NLP has the following types of ambiguities –
- **1. Lexical Ambiguity-** The ambiguity of a single word is called lexical ambiguity.
- For example, the word “back” can be a noun (back stage), an adjective (back door) or an adverb (back away).
- **2. Syntactic Ambiguity-** This kind of ambiguity occurs when a sentence is parsed in different ways.
- For example, the sentence “The man saw the girl with the telescope”. It is ambiguous whether the man saw the girl carrying a telescope or he saw her through his telescope.

- **3. Semantic Ambiguity-** This kind of ambiguity occurs when the meaning of the words themselves can be misinterpreted.
- In other words, semantic ambiguity happens when a sentence contains an ambiguous word or phrase.
- For example, the sentence “The car hit the pole while it was moving” is having semantic ambiguity because the interpretations can be “The car, while moving, hit the pole” and “The car hit the pole while the pole was moving”.
- **4. Anaphoric Ambiguity-** This kind of ambiguity arises due to the use of anaphora entities in discourse.
- For example, the horse ran up the hill. It was very steep. It soon got tired. Here, the anaphoric reference of “it” in two situations cause ambiguity.

- **5. Pragmatic ambiguity**-Such kind of ambiguity refers to the situation where the context of a phrase gives it multiple interpretations.
- In simple words, we can say that pragmatic ambiguity arises when the statement is not specific.
- For example, the sentence “I like you too” can have multiple interpretations like I like you (just like you like me), I like you (just like someone else dose).

