Bhilai Institute of Technology, Durg

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Natural Language Processing

UNIT 3: SEMANTIC ANALYSIS

Prepared By- Dr Shikha Pandey

UNIT - 3

Semantic analysis: Semantic Analysis: Meaning Representation, Lexical Semantics, Ambiguity, Word Sense Disambiguation. Discourse Processing: cohesion, Reference Resolution, Discourse Coherence and Structure. Knowledge Representation, reasoning.

Semantic Analysis: Meaning Representation

Semantic Analysis is a subfield of Natural Language Processing (NLP) that attempts to understand the meaning of Natural Language. Understanding Natural Language might seem a straightforward process to us as humans. However, due to the vast complexity and subjectivity involved in human language, interpreting it is quite a complicated task for machines. Semantic Analysis of Natural Language captures the meaning of the given text while taking into account context, logical structuring of sentences and grammar roles.

The semantic analysis method begins with a language-independent step of analyzing the set of words in the text to understand their meanings. This step is termed '**LEXICAL SEMANTICS**' and refers to fetching the dictionary definition for the words in the text. Subsequently, words or elements are parsed. Each element is designated a grammatical role, and the whole structure is processed to cut down on any confusion caused by ambiguous words having multiple meanings/

Parts of Semantic Analysis

Semantic Analysis of Natural Language can be classified into two broad parts:

1. Lexical Semantic Analysis: Lexical Semantic Analysis involves understanding the meaning of each word of the text individually. It basically refers to fetching the dictionary meaning that a word in the text is deputed to carry.

2. Compositional Semantics Analysis: Although knowing the meaning of each word of the text is essential, it is not sufficient to completely understand the meaning of the text.

For example, consider the following two sentences:-

Sentence 1: Students love BIT.

Sentence 2: BIT loves Students.

Although both these sentences 1 and 2 use the same set of root words {student, love, BIT}, they convey entirely different meanings. Hence, under Compositional Semantics Analysis, we try to understand how combinations of individual words form the meaning of the text. Example Sentences where semantic analysis required. (means how machine interpret these kind of sentence)

1. Elon musk is one of the co-founder of tesla, which is based in Austin,Tesla.

2. I got a blackberry.

3. I saw a man in the park with a telescope.

Tasks involved in Semantic Analysis

In order to understand the meaning of a sentence, the following are the major processes involved in Semantic Analysis:

1.Word Sense Disambiguation 2. Relationship Extraction

1. Word Sense Disambiguation:

We understand that words have different meanings based on the context of its usage in the sentence. If we talk about human languages, then they are ambiguous too because many words can be interpreted in multiple ways depending upon the context of their occurrence

In Natural Language, the meaning of a word may vary as per its usage in sentences and the context of the text. Word Sense Disambiguation involves interpreting the meaning of a word based upon the context of its occurrence in a text. In semantic analysis, word sense disambiguation refers to an automated process of determining the sense or meaning of the word in a given context. As natural language consists of words with several meanings (polysemic), the objective here is to recognize the correct meaning based on its use.

For example, ‘Raspberry Pi’ can refer to a fruit, a single-board computer, or even a company (UK-based foundation). Hence, it is critical to identify which meaning suits the word depending on its usage.For example, the word ‘Bark’ may mean ‘the sound made by a dog’ or ‘the outermost layer of a tree.’

Likewise, the word ‘rock’ may mean ‘a stone‘ or ‘a genre of music‘ – hence, the accurate meaning of the word is highly dependent upon its context and usage in the text.

Thus, the ability of a machine to overcome the ambiguity involved in identifying the meaning of a word based on its usage and context is called Word Sense Disambiguation.

Evaluation of WSD

The evaluation of WSD requires the following two inputs −

A Dictionary

The very first input for evaluation of WSD is dictionary, which is used to specify the senses to be disambiguated.

Test Corpus

Another input required by WSD is the high-annotated test corpus that has the target or correct-senses. The test corpora can be of two types;

Lexical sample − This kind of corpora is used in the system, where it is required to disambiguate a small sample of words.

All-words − This kind of corpora is used in the system, where it is expected to disambiguate all the words in a piece of running text.

Approaches and Methods to Word Sense Disambiguation (WSD)

Approaches and methods to WSD are classified according to the source of knowledge used in word disambiguation.

Let us now see the four conventional methods to WSD −

Dictionary-based or Knowledge-based Methods

As the name suggests, for disambiguation, these methods primarily rely on dictionaries, treasures and lexical knowledge base. They do not use corpora evidences for disambiguation.

Supervised Methods

For disambiguation, machine learning methods make use of sense-annotated corpora to train. These methods assume that the context can provide enough evidence on its own to disambiguate the sense. In these methods, the words knowledge and reasoning are deemed unnecessary. The context is represented as a set of “features” of the words. It includes the information about the surrounding words also. Support vector machine and memory-based learning are the most successful supervised learning approaches to WSD. These methods rely on substantial amount of manually sense-tagged corpora, which is very expensive to create.

Semi-supervised Methods

Due to the lack of training corpus, most of the word sense disambiguation algorithms use semi-supervised learning methods. It is because semi-supervised methods use both labelled as well as unlabeled data. These methods require very small amount of annotated

text and large amount of plain un annotated text. The technique that is used by semi supervised methods is bootstrapping from seed data.

Unsupervised Methods

These methods assume that similar senses occur in similar context. That is why the senses can be induced from text by clustering word occurrences by using some measure of similarity of the context. This task is called word sense induction or discrimination. Unsupervised methods have great potential to overcome the knowledge acquisition bottleneck due to non-dependency on manual efforts.

2. Relationship Extraction:

Relationship extraction is a procedure used to determine the semantic relationship between words in a text. In semantic analysis, relationships include various entities, such as an individual's name, place, company, designation, etc. Moreover, semantic categories such as, 'is the chairman of,' 'main branch located a'', 'stays at,' and others connect the above entities. Let's consider a phrase as an example. 'Elon Musk is one of the co-founders of Tesla, which is based in Austin, Texas.'

This phrase illustrates two different relationships.

Elon Musk is the co-founder of Tesla

[Person] [Company]

Tesla is based in Austin, Texas
[Company] [Place]

Elements of Semantic Analysis

Some of the critical elements of Semantic Analysis that must be scrutinized and taken into account while processing Natural Language are:

Hyponymy: Hyponymys refers to a term that is an instance of a generic term. They can be understood by taking class-object as an analogy. For example: ‘Color‘ is a hypernymy while ‘grey‘, ‘blue‘, ‘red‘, etc, are its hyponyms.

Homonymy: Homonymy refers to two or more lexical terms with the same spellings but completely distinct in meaning. For example: ‘Rose‘ might mean ‘the past form of rise‘ or ‘a flower‘, – same spelling but different meanings; hence, ‘rose‘ is a homonymy.

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).

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.

Meronomy: Meronomy refers to a relationship wherein one lexical term is a constituent of some larger entity. For example: ‘Wheel‘ is a meronym of ‘Automobile‘

Meaning Representation

While, as humans, it is pretty simple for us to understand the meaning of textual information, it is not so in the case of machines. Thus, machines tend to represent the text in specific formats in order to interpret its meaning. This formal structure that is used to understand the meaning of a text is called meaning representation.

Basic Units of Semantic System:

In order to accomplish Meaning Representation in Semantic Analysis, it is vital to understand the building units of such representations. The basic units of semantic systems are explained below:

Entity: An entity refers to a particular unit or individual in specific such as a person or a location. For example GeeksforGeeks, Delhi, etc.

Concept: A Concept may be understood as a generalization of entities. It refers to a broad class of individual units. For example Learning Portals, City, Students.

Relations: Relations help establish relationships between various entities and concepts. For example: 'GeeksforGeeks is a Learning Portal', 'Delhi is a City.', etc.

Predicate: Predicates represent the verb structures of the sentences.

In Meaning Representation, we employ these basic units to represent textual information.

Approaches to Meaning Representations:

Now that we are familiar with the basic understanding of Meaning Representations, here are some of the most popular approaches to meaning representation:

- First-order predicate logic (FOPL)
- Semantic Nets
- Frames
- Conceptual dependency (CD)
- Rule-based architecture
- Case Grammar
- Conceptual Graphs

Semantic Analysis Techniques

Based upon the end goal one is trying to accomplish, Semantic Analysis can be used in various ways. Two of the most common Semantic Analysis techniques are:

- ### Text Classification

  In-Text Classification, our aim is to label the text according to the insights we intend to gain from the textual data.

  For example:

  In Sentiment Analysis, we try to label the text with the prominent emotion they convey. It is highly beneficial when analyzing customer reviews for improvement.

  In Topic Classification, we try to categories our text into some predefined categories. For example: Identifying whether a research paper is of Physics, Chemistry or Maths

  In Intent Classification, we try to determine the intent behind a text message. For example: Identifying whether an e-mail received at customer care service is a query, complaint or request.

Text Extraction

In-Text Extraction, we aim at obtaining specific information from our text.
For Example,
In Keyword Extraction, we try to obtain the essential words that define the entire document.
In Entity Extraction, we try to obtain all the entities involved in a document.

Significance of Semantics Analysis

Semantics Analysis is a crucial part of Natural Language Processing (NLP). In the ever-expanding era of textual information, it is important for organizations to draw insights from such data to fuel businesses. Semantic Analysis helps machines interpret the meaning of texts and extract useful information, thus providing invaluable data while reducing manual efforts.

Besides, Semantics Analysis is also widely employed to facilitate the processes of automated answering systems such as chatbots – that answer user queries without any human interventions.

Applications of Semantic Analysis till 2022

1. Conversational chatbots:- These tools help resolve customer problems in minimal time, thereby increasing customer satisfaction.
2. Automated ticketing support:- It understands the text within each ticket, filters it based on the context, and directs the tickets to the right person or department (IT help desk, legal or sales department, etc.).
3. Sentiment analysis:- Several companies are using the **sentiment analysis** functionality to understand the voice of their customers, extract sentiments and emotions from text, and, in turn, derive actionable data from them. It helps capture the tone of customers when they post reviews and opinions on social media posts or company websites.
4. Search engine results:- Search engines use semantic analysis to understand better and analyze **user intent** as they search for information on the web. Moreover, with the ability to capture the context of user searches, the engine can provide accurate and relevant results.
5. Language translation:- Today, semantic analysis methods are extensively used by language translators.

DISCOURSE PROCESSING

The word "Discourse" in Hindi can be translated as "□□□□□-□□□□□□" (Vichar-Vimarsh) or "□□□□□□" (Pravachan), depending on the context.

Discourse integration is the fourth phase in NLP, and simply means contextualisation. Discourse integration is the analysis and identification of the larger context for any smaller part of natural language structure (e.g. a phrase, word or sentence).

During this phase, it's important to ensure that each phrase, word, and entity mentioned are mentioned within the appropriate context. This analysis involves considering not only sentence structure and semantics, but also sentence combination and meaning of the text as a whole.

Otherwise, when analyzing the structure of text, sentences are broken up and analyzed and also considered in the context of the sentences that precede and follow them, and the impact that they have on the structure of text. Some common tasks in this phase include: **information extraction, conversation analysis, text summarisation, discourse analysis**.

Here are some complexities of natural language understanding introduced during this phase:

- Understanding of the expressed motivations within the text, and its underlying meaning.
- Understanding of the relationships between entities and topics mentioned, thematic understanding, and interactions analysis.
- Understanding the social and historical context of entities mentioned.

Here are a few example sentences that can be analyzed in a discourse analysis study, focusing on different aspects such as coherence, cohesion, power dynamics, politeness strategies, and conversational implicature.

1. Coherence and Cohesion

Sentence: "*Despite the heavy rain, the match continued, and the players showed great determination.*"

- Analysis: The sentence maintains coherence through logical connectors ("despite") and cohesive devices ("and") that link ideas together, showing a cause-and-effect relationship and the progression of events.

2. Power Dynamics

Sentence: "*As your manager, I expect you to complete the report by end of day*."

- Analysis: This sentence illustrates a power dynamic where the speaker asserts authority ("As your manager") and uses a directive ("I expect") to communicate an expectation.

3. Politeness Strategies

Sentence: "*Would you mind closing the window? It's getting a bit chilly in here.*"

- Analysis: The sentence uses a politeness strategy by phrasing the request as a question ("Would you mind") and providing a reason ("It's getting a bit chilly"), making the request more indirect and polite.

4. Conversational Implicature

Sentence: "*I noticed you haven't submitted your assignment yet. The deadline is tomorrow."*

- Analysis: The speaker implies that the listener should submit the assignment soon without directly saying it, relying on the listener to infer the urgency from the context.

5. Identity and Social Roles

Sentence: "*As a mother, I believe it's important to teach children about responsibility.*"

- Analysis: The speaker's identity as a mother is highlighted to give weight to her opinion, emphasizing her role and the associated values.

6. Interpersonal Relations

Sentence: "*I'm really sorry, but I won't be able to attend the meeting.*"

- Analysis: The sentence shows an interpersonal relation where the speaker is using an apology ("I'm really sorry") to soften the impact of the negative news ("I won't be able to attend the meeting")

7. Turn-Taking

Dialogue:

- Person A: "*What did you think about the movie?*"

- Person B: "I thought it was fantastic! The storyline was gripping and the acting superb."

- Analysis: This exchange demonstrates smooth turn-taking, where Person A's question prompts Person B's detailed response, maintaining the flow of conversation.

8. Topic Management

Dialogue:

- Person A: *"We should discuss the budget for the new project."*

- Person B: "A*bsolutely. Speaking of which, have you seen the latest financial reports?"*

- Analysis: Person B skillfully shifts the topic from a general discussion about the budget to the specific financial reports, indicating a seamless transition in topic management.

9. Speech Acts

Sentence: *"I hereby declare the meeting adjourned."*

- Analysis: This sentence is a performative speech act where the speaker is performing an action (adjourning the meeting) through their utterance.

These examples provide a range of contexts and features that can be explored in discourse analysis to understand how language functions in different social and communicative situations.

If we talk about the major problems in NLP, then one of the major problems in NLP is discourse processing − building theories and models of how utterances stick together to form **coherent discourse**. Actually, the language always consists of collocated, structured and coherent(consistent, or logically connected) groups of sentences rather than isolated and unrelated sentences like movies. These coherent groups of sentences are referred to as discourse.

Concept of Coherence

Coherence and discourse structure are interconnected in many ways. Coherence, along with property of good text, is used to evaluate the output quality of natural language generation system. The question that arises here is what does it mean for a text to be coherent? Suppose we collected one sentence from every

page of the newspaper, then will it be a discourse? Of-course, not. It is because these sentences do not exhibit coherence. The coherent discourse must possess the following properties −

Coherence relation between utterances

The discourse would be coherent if it has meaningful connections between its utterances. This property is called coherence relation. For example, some sort of explanation must be there to justify the connection between utterances.

Relationship between entities

Another property that makes a discourse coherent is that there must be a certain kind of relationship with the entities. Such kind of coherence is called entity-based coherence.

Discourse structure

An important question regarding discourse is what kind of structure the discourse must have. The answer to this question depends upon the segmentation we applied on discourse. Discourse segmentations may be defined as determining the types of structures for large discourse. It is quite difficult to implement discourse segmentation, but it is very important for information retrieval, text summarization and information extraction kind of applications.

Text Coherence

Lexical repetition is a way to find the structure in a discourse, but it does not satisfy the requirement of being coherent discourse. To achieve the coherent discourse, we must focus on coherence relations in specific. As we know that coherence relation defines the possible connection between utterances in a discourse. Hebb has proposed such kind of relations as follows −
We are taking two terms S0 and S1 to represent the meaning of the two related sentences −

Result

It infers that the state asserted by term S0 could cause the state asserted by S1. For example, two statements show the relationship result: Ram was caught in the fire. His skin burned.

Explanation

It infers that the state asserted by S1 could cause the state asserted by S0. For example, two statements show the relationship − Ram fought with Shyam's friend. He was drunk.

Parallel

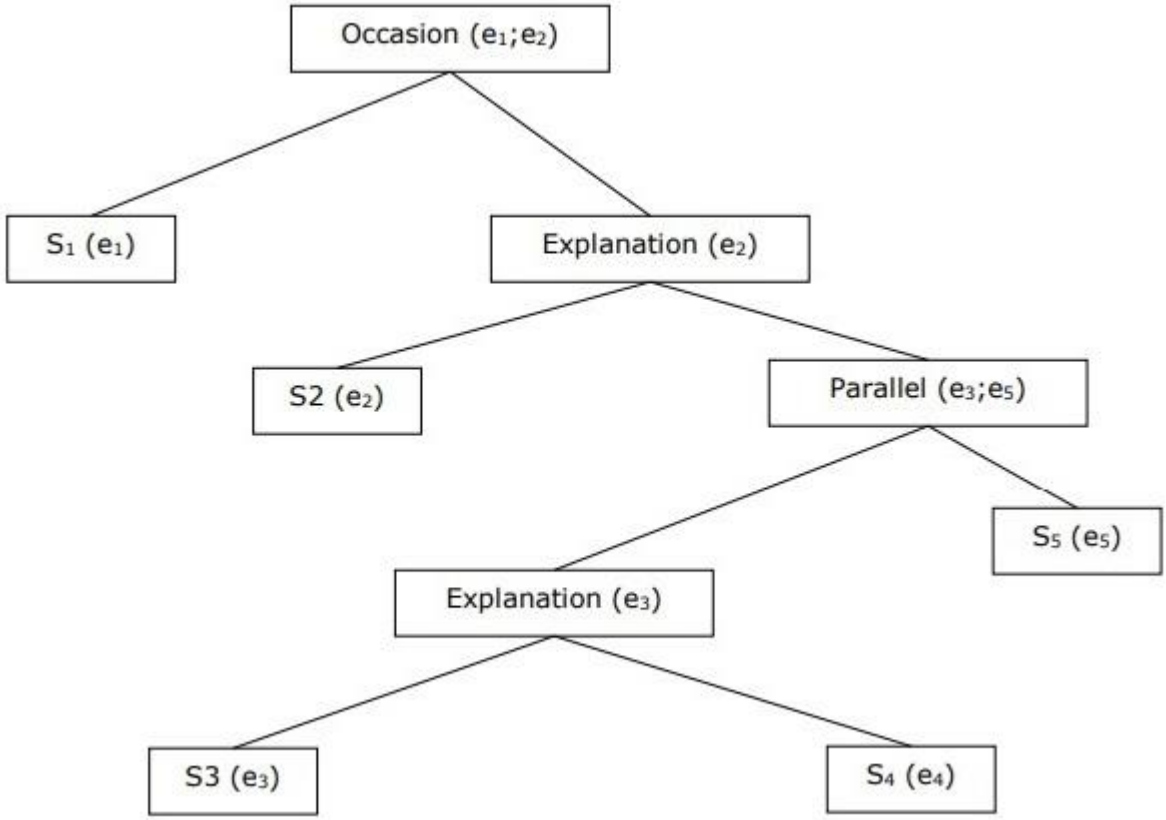
It infers p(a1,a2,…) from assertion of S0 and p(b1,b2,…) from assertion S1. Here ai and bi are similar for all i. For example, two statements are parallel − Ram wanted car. Shyam wanted money.
Elaboration
It infers the same proposition P from both the assertions − S0 and S1 For example, two statements show the relation elaboration: Ram was from Chandigarh. Shyam was from Kerala.

Occasion

It happens when a change of state can be inferred from the assertion of S0, final state of which can be inferred from S1 and vice-versa. For example, the two statements show the relation occasion: Ram picked up the book. He gave it to Shyam.

Building Hierarchical Discourse Structure

The coherence of entire discourse can also be considered by hierarchical structure between coherence relations. For example, the following passage can be represented as hierarchical structure −
S1 − Ram went to the bank to deposit money.
S2 − He then took a train to Shyam's cloth shop.
S3 − He wanted to buy some clothes.
S4 − He do not have new clothes for party.
S5 − He also wanted to talk to Shyam regarding his health



Reference Resolution

Interpretation of the sentences from any discourse is another important task and to achieve this we need to know who or what entity is being talked about. Here, interpretation reference is the key element. **Reference** may be defined as the linguistic expression to denote an entity or individual. For example, in the passage, Ram, the manager of ABC bank, saw his friend Shyam at a shop. He went to meet him, the linguistic expressions like Ram, His, He are reference.

On the same note, **reference resolution** may be defined as the task of determining what entities are referred to by which linguistic expression.

Terminology Used in Reference Resolution

We use the following terminologies in reference resolution −

- **Referring expression** − The natural language expression that is used to perform reference is called a referring expression. For example, the passage used above is a referring expression.

- **Referent** − It is the entity that is referred. For example, in the last given example Ram is a referent.
- **Corefer** − When two expressions are used to refer to the same entity, they are called corefers. For example, ***Ram*** and ***he*** are corefers.
- **Antecedent** − The term has the license to use another term. For example, ***Ram*** is the antecedent of the reference ***he***.
- **Anaphora & Anaphoric** − It may be defined as the reference to an entity that has been previously introduced into the sentence. And, the referring expression is called anaphoric.
- **Discourse model** − The model that contains the representations of the entities that have been referred to in the discourse and the relationship they are engaged in.

Types of Referring Expressions

Let us now see the different types of referring expressions. The five types of referring expressions are described below −

Indefinite Noun Phrases

Such kind of reference represents the entities that are new to the hearer into the discourse context. For example − in the sentence Ram had gone around one day to bring him some food − some is an indefinite reference.

Definite Noun Phrases

Opposite to above, such kind of reference represents the entities that are not new or identifiable to the hearer into the discourse context. For example, in the sentence - I used to read The Times of India – The Times of India is a definite reference.

Pronouns

It is a form of definite reference. For example, Ram laughed as loud as he could. The word he represents pronoun referring expression.

Demonstratives

These demonstrate and behave differently than simple definite pronouns. For example, this and that are demonstrative pronouns.

Names

It is the simplest type of referring expression. It can be the name of a person, organization and location also. For example, in the above examples, Ram is the name-refereeing expression.

Reference Resolution Tasks

The two reference resolution tasks are described below.

Coreference Resolution

It is the task of finding referring expressions in a text that refer to the same entity. In simple words, it is the task of finding corefer expressions. A set of coreferring expressions are called coreference chain. For example - He, Chief Manager and His - these are referring expressions in the first passage given as example.

Constraint on Coreference Resolution

In English, the main problem for coreference resolution is the pronoun it. The reason behind this is that the pronoun it has many uses. For example, it can refer much like he and she. The pronoun it also refers to the things that do not refer to specific things. For example, It's raining. It is really good.

Pronominal Anaphora Resolution

Unlike the coreference resolution, pronominal anaphora resolution may be defined as the task of finding the antecedent for a single pronoun. For example, the pronoun is his and the task of pronominal anaphora resolution is to find the word Ram because Ram is the antecedent.

Questions

1. What are the different approaches to meaning representation in natural language processing (NLP), and how do they differ in handling complex sentences?
2. How does compositional semantics contribute to the construction of meaning representation in a sentence?
3. Explain the role of predicate logic in meaning representation. How does it help in formalizing the meaning of sentences?
4. What is the difference between homonymy and polysemy in lexical semantics? Provide examples of each.
5. How do semantic networks represent lexical relationships, and what are their applications in NLP?
6. Discuss the role of thesauri and ontologies in capturing lexical semantics. How do they differ from each other?
7. What are the different types of ambiguity in natural language? Provide examples for each.
8. Discuss the role of context in word sense disambiguation. What are some common techniques used to capture contextual information?
9. Compare and contrast knowledge-based and machine learning approaches to word sense disambiguation. What are the advantages and limitations of each? What is the role of cohesion in discourse processing, and how does it contribute to the overall coherence of a text?
10. Explain the difference between lexical cohesion and grammatical cohesion with examples.
11. How do anaphoric references contribute to text cohesion? Provide examples of how they are resolved in discourse.
12. What are the different types of reference resolution in discourse processing? Provide examples of each.
13. How does the process of co-reference resolution work in natural language processing, and what are its main challenges?
14. Discuss the role of machine learning in improving reference resolution techniques. What are some commonly used algorithms?
15. What is the difference between local and global coherence in discourse processing? Provide examples to illustrate your answer.
16. How does the Rhetorical Structure Theory (RST) help in understanding the structure and coherence of discourse?
17. Discuss the role of discourse markers in maintaining coherence in a conversation or written text. Provide examples of how they function.
18. What are the main types of knowledge representation in artificial intelligence, and how do they differ from each other?
19. Explain the role of frames and scripts in representing knowledge. How do they help in understanding natural language?
20. What are the different types of reasoning used in artificial intelligence, and how do they apply to natural language processing?
21. How does logical reasoning differ from probabilistic reasoning in the context of NLP?
22. Discuss the role of reasoning in question-answering systems. How does it contribute to generating accurate and relevant responses?