Q12.(a) What is NLP? Discuss various Natural Language Processing techniques.

MDU BCA Dec. 2015, April 2015

Ans. Natural Language Processing (NLP):-

Natural language processing is a field of artificial intelligence and linguistics. It focuses on developing systems that allow computers to communicate with people using everyday language.

OR

Natural language processing (NLP) can be defined as the automatic (or semi-automatic) processing of human language.

OR

Natural language processing is the engineering of systems that processes or analyzes written or spoken natural language.

Natural language processing is a significant area of AI because a computer would be considered intelligent if it can understand the commands given in natural language.

The main task of natural language processing is to deal with the interactions between computers and human languages. This in itself is an entire separate area of computer science, which is called human-computer interaction.

The main problems that face people working on natural language processing are summed up in the task of natural language understanding, which means enabling

computers to understand in a certain way what human language input is meant to convey.

The most recent natural language processing ideas involve machine learning, and more specifically statistical machine learning. Machine learning is also a branch of artificial intelligence and it concerns the construction and study of systems that can learn from data.

Some of the most common topics of research in natural language processing include: machine translation, which involves translating the input text from one human language to another. This is one of the hardest problems and requires a very wide range of knowledge types in order for it to be solved.

There is also automatic summarization, which involves producing a readable summary of a passage or a text. This is an often used application of natural language processing. There is discourse analysis, and it includes a lot of tasks such as identifying the discourse structure of text, and recognizing and classifying speech acts in a chunk of text. Other commonly known applications are: conference resolution, name entity recognition, natural language generation, natural language understanding, optical character recognition, question answering, speech recognition and sentiment analysis.

This is just a small portion of what natural language processing involves. And, the span of what natural language processing can become in the future is very wide. One of the most difficult problems still facing

professionals in the field is human-level natural language processing which if solved is equivalent to solving the central artificial intelligence problem that is making computers as intelligent as people.

The future of natural language processing is therefore tied closely to the development of artificial intelligence. As natural language understanding improves, future computers will have the ability to obtain data and learn online and apply that in the real world. Combined with natural language generation, computers will soon be more capable taking in and giving out instructions.

Goal of NLP:-

The goal of NLP is to accomplish human-like language processing i.e. the goal of NLP is to design and build a computer system that will analyze, understand, and generate natural human-languages.

A complete NLP system would be able to:

- ***Translate of one human-language text to another.***
- ***Generate human-language text such as fiction, manuals, and general descriptions.***
- ***Interface to other systems such as databases and robotic systems thus enabling the use of human-language type commands and queries.***
- ***Answer questions about the contents of the text.***
- ***Understand human-language text to provide a summary or to draw conclusions.***

There are more practical goals for NLP, many related to the particular application for which it is being utilized.

Natural Language Processing techniques:

Following are the various techniques for natural language processing are:

1. Pattern Matching
2. Syntactically-driven Parsing
3. Semantic Grammars
4. Case frame instantiation
5. Robust Parsing

1. Pattern Matching:

This technique involves interpreting input utterances as a whole, rather than building up their interpretation by combining the structure and meaning of words or other lower-level constituents. The approach is thus wholistic rather than constructive. With this approach, the interpretations are obtained by matching patterns of words against the input utterance. Associated with each pattern is an interpretation, so that the derived interpretation is the one attached to the pattern that matched.

2. Syntactically-driven Parsing:

This technique deals with the ways that words can fit together to form higher level units such as phrases, clauses and sentences. Syntactically driven parsing means interpretation of larger groups of words are built up out of

the interpretation of their syntactic constituent words or phrases. In a way this is the opposite of pattern matching as here the interpretation of the input is done as a whole.

3. Semantic Grammars:

Natural language analysis based on semantic grammar is similar to syntactically driven parsing except that in semantic grammar the categories used are defined semantically and syntactically. Thus, here semantic grammar is also involved. However, this technique only works properly in restricted domains. Thus, it is a technique useful only for applied natural language processing, not for general NLP.

4. Case frame instantiation:

Case frame instantiation is one of the major parsing techniques under active research today. It has some very useful computational properties such as its recursive nature and its ability to combine bottom-up recognition of key constituents with top-down instantiation of less structured constituents.

5. Robust Parsing:

Any natural language interface which is used in a practical application with a multitude of users must be able to handle input that is outside its grammar or expectations in various ways. Methods of robust parsing are under active investigation at the moment with the chief outstanding problem being the coordination of multiple, independent, construction-specific parsing strategies on the same input.

Q12.(c) What are the main stages of Natural Language Analysis? Explain.

OR

What are the major aspects of any natural language understanding theory? Explain.

Ans. There are three main stages of natural language analysis. These are:

1. Syntactic Analysis
2. Semantic Analysis
3. Pragmatic Analysis

1. Syntactic Analysis:-

In this stage grammatical rules are used describing the legal structure of the language to obtain one or more parses of the sentence.

The input utterance is being checked to ensure that its syntax is correct and structured representations of the possible parses are generated.

The output of this level of processing is the representation of the sentence that reveals the structured dependency relationships between the words.

For example, consider two sentences: "The dog chased the cat." And "The cat chased the dog". These two sentences convey different meanings.

Some word sequences may be rejected if they violate the rules of the language for how words may be combined. For example, "the cat the dog chased".

2. Semantic Analysis:-

Semantic analysis deals with the meaning of the sentence. In this analysis, it is tried to capture the meaning conveyed by a sentence. The structures created by the syntactic analyzer are assigned meaning. Thus, a mapping is made between the syntactic structures and the objects in the task domain. The structures for which no such mapping is possible are rejected.

3. Pragmatic Analysis:-

The pragmatic component explains how the utterances relate to the world.

This is a high-level knowledge which relates the use of sentences in different contexts and how the context affects the meaning of the sentence. Thus, in this stage the initial representations are analyzed based on contextual and world knowledge and a final meaning representation is obtained. Additional contextual information is used to fill in gaps in the meaning representation, and to work out what the speaker was really getting at.

To understand the difference among these aspects, consider the following sentences, which might appear at the start of an AI textbook:

- This book is about artificial intelligence.
- The green frogs sleep soundly.
- Colorless green ideas sleep furiously.
- Furiously sleep ideas green colorless.

The first sentence would be quite appropriate at the start of such a book; it is syntactically, semantically, and pragmatically well formed. The second sentence is syntactically and semantically well formed, but it would appear very strange at the start of an AI book; it is thus not pragmatically well formed for that context. The third sentence is syntactically well formed, but it is semantically non-sensical. The fourth sentence is syntactically ill formed; it does not make any sense - syntactically, semantically, or pragmatically.

Q12.(e) Explain pragmatic processing.

MDU BCA Nov. 2016, May 2016

Ans. Pragmatic Processing

Pragmatic Processing is the last stage of Natural Language Processing. It refers to the contextual interpretation.

The pragmatic component explains how utterances relate to the world. In this process, meaning of words is elaborated based on the contextual and world knowledge and a final meaning representation is obtained.

This is high-level knowledge which relates the use of sentences in different contexts and how the context affects the meaning of the sentence.

To understand the concept, consider the following sentences which might appear at the start of an AI textbook:

- This book is about artificial intelligence.
- The green frogs sleep soundly.
- Colorless green ideas sleep furiously.
- Furiously sleep ideas green colorless.

The first sentence would be quite appropriate at the start of such a book; it is syntactically, semantically, and pragmatically well formed.

The second sentence is syntactically and semantically well formed, but it would appear very strange at the start

of an AI book; it is thus not pragmatically well formed for that context.

The third sentence is syntactically well formed, but it is semantically non-sensical.

The fourth sentence is syntactically ill formed; it does not make any sense - syntactically, semantically, or pragmatically.

Q13.(a) Define syntactic processing. How is it used and useful? Explain with examples.

MDU BCA Nov. 2016

OR

How is syntactic processing done? Explain the two basic top-down and bottom-up parsing techniques used for syntactic processing.

Ans. Syntax provides rules to put together words to form components of sentence and to put together these components to form sentences. Parsing is a method to perform syntactic analysis of a sentence.

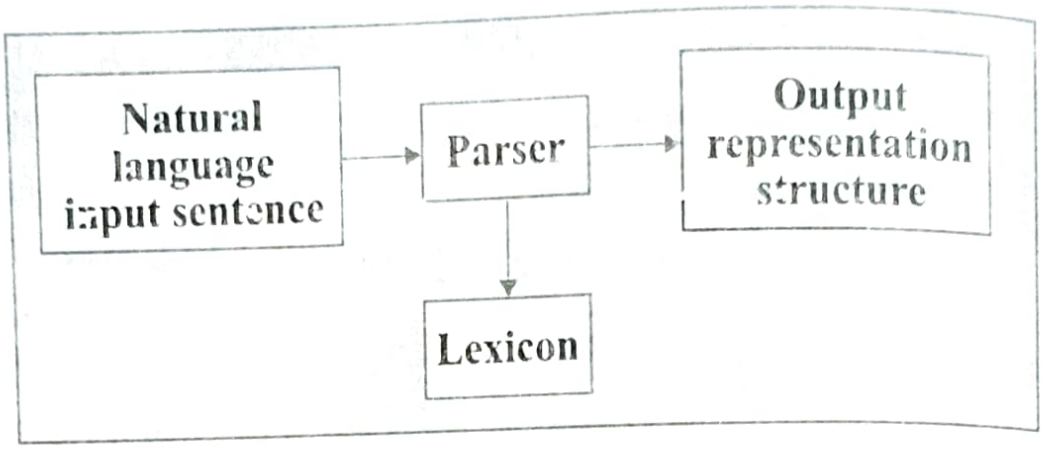
To perform the syntactic processing, the knowledge of grammar and parsing technique is required. The grammar is formal specification of rules allowable in the language and parsing is a method of analyzing a sentence to determine its structure according to grammar.

Syntactic processing is done using parsing. The parser generates basic syntactic structure of the sentence. This stage uses grammatical information to perform some structural preprocessing on the input. It performs the task of applying grammar rules and computes syntactic representation of the meaning. This preprocessing stage is called parsing.

The parsing performs grouping and labeling of parts of a sentence in a way that displays their relationships to each other in a useful way. It checks the validity of a sentence. Hence a sentence, if grammatically

correct and semantically incorrect will be considered as correct after first phase.

The basic parsing technique is shown below:-



The parser is a computer program, which accepts the natural language sentence as input and generates an output structure suitable for analysis. The lexicon is a dictionary of words, where each word contains some syntactic, some semantic and possibly some pragmatic information.

Two basic parsing techniques are: top-down parsing and bottom-up parsing.

Top-down parsing:-

A simple parsing algorithm used is the top-down parsing algorithm. In top-down parsing, grammar rules are applied to a starting symbol (i.e. sentence) until it generates terminal symbols, and the terminal symbols are found in the object sentence. Since all non-terminal symbols are functional blocks, any non-terminal symbol may be viewed as a starting symbol. When all of the terminal symbols are found, and no unknown terminal

symbols remain in the sentence, then the top-down parsing succeeds.

Thus, in top-down parsing, words of sentence are replaced by their categories like noun phrase (NP), verb phrase (VP) etc. These symbols like (NP, VP) are rewritten as per grammar rules. Finally, terminal symbols are replaced by language words.

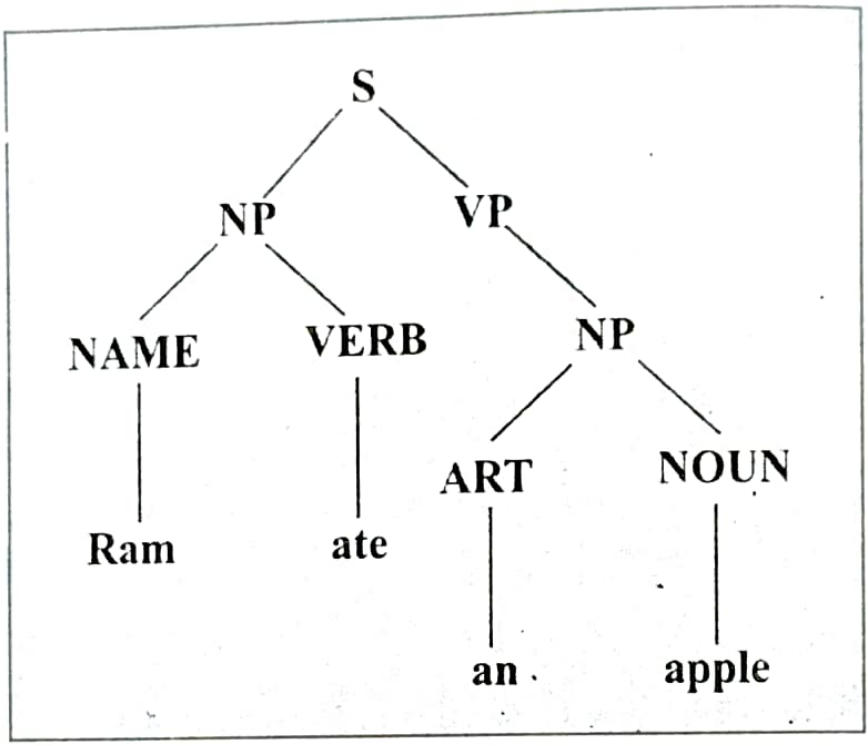
To illustrate the concept, consider the following example: "Ram ate an apple"

This consists of a noun phrase 'Ram' and a verb phrase 'ate an apple'. The noun phrase is just a proper noun, while the verb phrase consists of a verb 'ate' and another noun phrase 'an apple'. This noun phrase consists of a determiner 'an' and a noun 'apple'. The complete top-down parse and parsing tree for this is given below:

S → NP VP	(Replace S by its linguistic constituents NPVP)
→ Name VP	(Replace NP by Name)
→ Ram VP	(Replace constituent category name by terminal value)
→ Ram Verb NP	(Replace VP by constituent category)
→ Ram ate NP	(Replace verb by terminal value)
→ Ram ate ART noun	(Replace NP by category)

→ Ram ate an noun (Replace ART by terminal value)

→ Ram ate an apple (Replace noun by terminal value)



Bottom-up parsing:-

Bottom-up parsing is more general than top-down parsing. In Bottom-up parsing we start with the sentence and try to apply the production rules in reverse, in order to finish up with the start symbol of the grammar. This corresponds to starting at the leaves of the parse tree, and working back to the root. The bottom up parse for sentence "Ram ate an apple" would be as follows:

→ Ram ate an apple.

→ Name ate an apple. (replace Ram)

→ Name verb an apple. (replace ate)

→ Name verb art apple. (replace an)

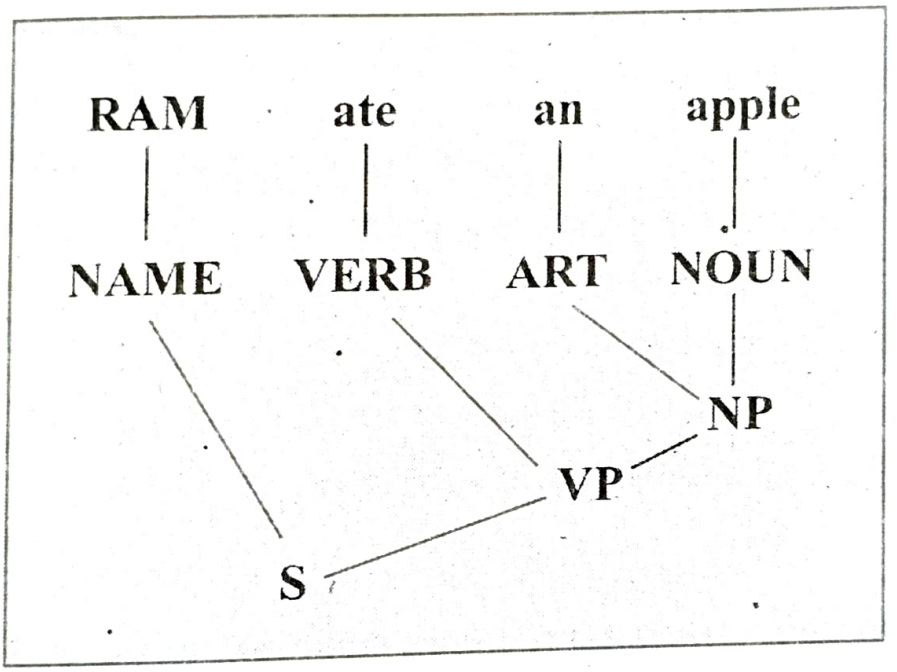
→ NP verb art apple. (replace Name)

→ NP verb NP (replace Noun phrase)

→ NP VP (replace verb NP)

→ S (replace NP VP)

The parse tree of the above sentence is shown below:-



The incorrect sentences will be excluded by the grammar. For example, 'Apple Ram ate' starts with two nouns, which is not allowed in the grammar. However, some odd sentences will be allowed, such as 'an apple ate Ram'. This sentence is syntactically acceptable, just semantically odd, so should still be parsed. Note that sentences, which are syntactically sound but make no sense, are acceptable by the parser. Thus, the next stage (semantic processing) would be needed to filter out these nonsensical parses.

Q13.(b) What is Semantic Processing? What are the approaches to semantic analysis? Explain. Also differentiate between Syntactic, Semantic and Pragmatic processing.

Ans. Semantic Processing:-

Semantics is the study of the meaning of words and semantic analysis is the analysis used to extract meaning from utterances.

Semantic processing determines the possible meanings of a sentence by focusing on the interactions among word-level meanings in the sentence. The structures created by the syntactic analyzer are assigned meaning. Thus, a mapping is made between the syntactic structures and the objects in the task domain. The structures for which no such mapping is possible are rejected.

For example: the sentence "There are colourless red ideas" would be rejected because colourless and red make no sense although the sentence is correct.

Approaches to semantic analysis:-

Following are the approaches to semantic analysis:-

1. Syntax driven semantic analysis approach
2. Lexical semantic approach
3. Compositional semantic approach

1. **Syntax driven semantic approach:-**

This is the elementary approach of semantic analysis though its scope is limited. This approach assigns meaning representation to input, based on the knowledge from the lexicon (dictionary) and the grammar. The basic idea behind this approach is that the meaning of a sentence can be composed from the meaning of its parts. This idea is not very effective because according to this principle, to find out the meaning of a sentence, the basic word meaning will play prime role but it does not include the ordering of words and relations between them in the sentence.

2. **Lexical semantic approach:-**

This approach makes the use of semantic grammar. Using this approach, input sentences are transformed through the domain dependent semantic rules.

Thus, it uses conceptual dependency theory. Conceptual dependency structures provide a form of linked knowledge that can be used in larger structures.

3. **Compositional semantic approach:-**

In compositional semantic approach, the meaning of an expression is derived from the meaning of the parts of that expression. Here the target knowledge structures constructed are typically logical expressions like the formulas of FOPL.

Following are the differences between Syntactic, Semantic and Pragmatic processing:

Syntactic Processing	Semantic Processing	Pragmatic Processing
1. In syntactic processing, sentences are checked according to the grammar.	1. In semantic processing, sentences are checked whether they are meaningful or not.	1. In pragmatic processing, sentences are checked whether they are true or false.
2. In this processing, grammar and language rules are used.	2. In this processing, semantic grammar is used.	2. In this processing, high level knowledge base is required.
3. Parsing structure is used for checking the sentence.	3. Lexical analyzer, case grammar and semantic grammar are used for checking the meaning of the sentence.	3. Knowledge database is used to check the structure of sentence (behaviour).
4. If a sentence has syntax anomaly, it will be rejected. e.g. Dog the bite a man.	4. If a sentence has semantic anomaly, it will be rejected. e.g. Colourless green ideas sleep furiously.	4. If a sentence has pragmatic anomaly, it will be rejected. e.g. Washington was 5th president of U.S.A.
5. In any natural language understanding, it is the first step.	5. It is the next step after the syntactic processing.	5. It is the final step after syntactic and semantic processing.