**Chapter 13**

**Natural language processing**

**The chapter consist of Short type Questions &Answers , Descriptive Question & Answer and MCQs & answers.**

Contents

[Short type Questions &Answers 1](#_Toc14943864)

[Q What is NLP 2](#_Toc14943865)

[Q List some components of NLP 2](#_Toc14943866)

[Q Define NLP Terminology. 2](#_Toc14943867)

[Q What is tokenization ? 2](#_Toc14943868)

[QWhat is stemming ? 2](#_Toc14943869)

[QWhat is lemmatizing 2](#_Toc14943870)

[Q What is Normalization ? 3](#_Toc14943871)

[Q What is POS (parts of speech) tagging ? 3](#_Toc14943872)

[Q What is NER (name entity recognition)? 3](#_Toc14943873)

[What are nlp libraries and tools ? 3](#_Toc14943874)

[Q List some areas of NLP 3](#_Toc14943875)

[Descriptive Question & Answer 4](#_Toc14943876)

[Q What is part of Speech (POS)Tagging ? 4](#_Toc14943877)

[Q In the case of processing natural language, we normally mentioned one common terminology NLP and binding every language with the same terminology properly. Please explain in details about this NLP terminology with an example? 4](#_Toc14943878)

[Q Explain details about varieties areas available in case of processing natural languages smartly, whether we know impacted areas are very small as this processing starting very recently? 5](#_Toc14943879)

[Q Explain Lexical Analysis 5](#_Toc14943880)

[Q Explain Goals, Applications and Examples of Natural Language Processing(NLP) 7](#_Toc14943881)

[MCQs & answers 11](#_Toc14943882)

# Short type Questions &Answers

## Q What is NLP

Natural Language Processing or **NLP** is an automated way to understand or analyze the natural languages and extract required information from such data by applying machine learning Algorithms.

## Q List some components of NLP

Below are the few major components of NLP.

* Entity extraction: It involves segmenting a sentence to identify and extract entities, such as a person (real or fictional), organization, geographies, events, etc.
* Syntactic analysis: It refers to the proper ordering of words.
* Pragmatic analysis: Pragmatic Analysis is part of the process of extracting information from text.

## Q Define NLP Terminology.

NLP Terminology is based on the following factors:

* **Weights and Vectors:** TF-IDF, length(TF-IDF, doc), Word Vectors, Google Word Vectors
* **Text Structure:** Part-Of-Speech Tagging, Head of sentence, Named entities
* **Sentiment Analysis:** Sentiment Dictionary, Sentiment Entities, Sentiment Features
* **Text Classification:** Supervised Learning, Train Set, Dev(=Validation) Set, Test Set, Text Features, LDA.
* **Machine Reading:** Entity Extraction, Entity Linking,dbpedia, FRED (lib) / Pikes

|  |
| --- |
| Q What is tokenization ? |
| Splitting the sentence into words |
| QWhat is stemming ? |
| Stemming is the process of reducing a word to its word stem that affixes to suffixes and prefixes. |
| QWhat is lemmatizing **?** |
| Lemmatizing is also same like stemming but the difference is lemmantizing words known with dictionary. |
| Q What is Normalization ? |
| Converting different range of values to same scale from 0 to 1. |
| Q What is POS (parts of speech) tagging ? |
| Tagging a word with noun, pronoun, adverd, adjective etc. |
| Q What is NER (name entity recognition)? |
| NER refers to name entiyy recognization like places, organizations, companies etc. |
| What are nlp libraries and tools ? |
|  |
| CoreNLP from Stanford group. |
| NLTK, the most widely-mentioned NLP library for Python. |
| TextBlob, a user-friendly and intuitive NLTK interface. |
| Gensim, a library for document similarity analysis. |
| SpaCy, an industrial-strength NLP library built for performance. |

## Q List some areas of NLP

Natural Language Processing can be used for

* Semantic Analysis
* Automatic summarization
* Text classification
* Question Answering

Some real-life example of NLP is **IOS Siri**, the **Google assistant**, **Amazon echo**.

# Descriptive Question & Answer

## Q What is part of Speech (POS)Tagging ?

According to The Stanford Natural Language Processing Group :

A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc.

PoS taggers use an algorithm to label terms in text bodies. These taggers make more complex categories than those defined as basic PoS, with tags such as “noun-plural” or even more complex labels. Part-of-speech categorization is taught to school-age children in English grammar, where children perform basic PoS tagging as part of their education.

## Q In the case of processing natural language, we normally mentioned one common terminology NLP and binding every language with the same terminology properly. Please explain in details about this NLP terminology with an example?

**Answer:**  
This is the basic NLP Interview Questions asked in an interview. There have some several factors available in case of explaining natural language processing. Some of the key factors are given below:

* **Vectors and Weights**: Google Word vectors, length of TF-IDF, varieties documents, word vectors, TF-IDF.
* **Structure of Text**: Named Entities, tagging of part of speech, identifying the head of the sentence.
* **Analysis of sentiment**: Know about the features of sentiment, entities available for the sentiment, sentiment common dictionary.
* **Classification of Text**: Learning supervising, set off a train, set of validation in Dev, Set of define test, a feature of the individual text, LDA.
* **Reading of Machine Language**: Extraction of the possible entity, linking with an individual entity, DBpedia, some libraries like Pikes or FRED.

## Q Explain details about varieties areas available in case of processing natural languages smartly, whether we know impacted areas are very small as this processing starting very recently?

**Answer:**  
Natural language processing (NLP) can have an implementation in various areas of the current industry environment. Some of the key areas are explaining below:

* An analysis was done in a semantic way.
* Summarize natural language information automatically.
* Classification of varieties text is written in natural language.
* Ready answer to some common questions

We can give some key example of real life where natural language processing (NLP) used broadly. Examples are Google Assistance, IOS Siri, or Amazon echo.

## Q Explain Lexical Analysis

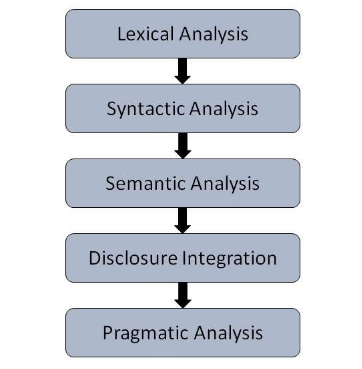
Lexical Analysis − It involves distinguishing and analyzing the structure of words.

Lexicon of a language means that the gathering of words and phrases in a verylanguage.

Lexical analysis is dividing the whole chunk of txt into paragraphs, sentences, and words.

Syntactic Analysis (Parsing) − It involves analysis of words within the sentence for synchronic linguistics and composing words in a very manner that shows the linkamong the words.

The sentence like “The faculty goes to boy” is rejected by English grammar analyser.



Semantic Analysis − It attracts the precise which means or the wordbook which means from the text.

The text is checked for meaningfulness.

It is done by mapping syntactical structures and objects within the task domain.

The linguistics analyser disregards sentence like “hot ice-cream”.

Discourse Integration − The which means of any sentence depends upon the which means of the sentence simply before it.

In addition, it additionally brings concerning the which means of like a shotsucceeding sentence.

Pragmatic Analysis − throughout this, what was aforesaid is re-interpreted on what it truly meant.

It involves deriving those aspects of language which require real world knowledge.

Implementation Aspects of Syntactic Analysis

There are a number of algorithms researchers have developed for syntactic analysis, but we consider only the following simple methods −

* Context-Free Grammar
* Top-Down Parser

## Q Explain Goals, Applications and Examples of Natural Language Processing(NLP)

**GOALS OF NLP**

The goal of natural language processing is to specify a language comprehension and production theory to such a level of detail that a person is able to write a computer program which can understand and produce natural language. The basic goal of NLP is to accomplish human like language processing. The choice of word “processing” is very deliberate and should not be replaced with “understanding”. For although the field of NLP was originally referred to as Natural Language Understanding (NLU), that goal has not yet been accomplished. A full NLU system would be able to:

      Paraphrase an input text.

   Translate the text into another language.

   Answer questions about the contents of the text.

   Draw inferences from the text.

While NLP has made serious inroads into accomplishing goals from first to third, the fact that NLP system can not, of themselves, draw inferences from text, NLU still remains the goal of NLP. Also there are some practical applications of NLP. An NLP-based IR system has the goal of providing more precise, complete information in response to a user’s real information need. The goal of the NLP system is to represent the true meaning and intent of the user’s query, which can be expressed as naturally in everyday language.

**APPLICATIONS OF NLP**

NLP lie in a number of disciplines like computer and information sciences, linguistics, mathematics, electrical and electronic engineering, artificial intelligence and robotics, psychology etc. Applications of NLP include a number of fields of studies such as machine translation, natural language text processing, summarization, user interfaces multilingual and Gross language information retrieval (CLIR), speech recognition, artificial intelligence and expert system. Research on NLP is regularly published in a number of conferences such as the annual proceedings of ACL (Association of Computational Linguistics) and its European counter part EACL, biennial proceedings of the Message Understanding Conferences (MUCS), Text Retrieval Conferences (TRECS) and ACM-SIGIR (Association of Computing Machinery-Special Interest Group on Information Retrieval) conferences.

As natural language processing technology matures, it is increasingly being used to support other computer applications. Such use naturally falls into two areas, one in which linguistic analysis merely serves as an interface to the primary program and the second one in which natural language considerations are central to the application. Natural language interfaces into a request in a formal database query language, and the program then proceeds as it would without the use of natural language processing techniques. The design of question answering systems is similar to that for interfaces to database management systems. One difference however, is that the knowledge base supporting the question answering system does not have the structure of a database. Similarly in message understanding systems, a fairly complete linguistic analysis may be required but the messages are relatively short and the domain is often limited. Also some more application areas include information and text categorization. In both applications, natural language processing imposes a linguistic representation on each document being considered. In text categorization a collection of documents is inspected and all documents are grouped into several categories based on the characteristics of the linguistic representations of the documents. In information filtering documents satisfying some criterion are singled out from a collection.

**Discourse Knowledge**

While syntax and semantics work with sentence-length units, the discourse level of NLP works with units of text longer than a sentence i.e. it does not interpret multi-sentence texts as just concatenated sentences, each of which can be interpreted singly. Discourse focuses on the properties of the text as a whole that convey meaning by making connections between component sentences. Several types of discourse processing can occur at this level like anaphora resolution and discourse/text structure recognition. Anaphora resolution is the replacing of words such as pronouns which are semantically vacant with the appropriate entity to which they refer. For example, newspaper articles can be deconstructed into discourse components such as: lead, main story, previous events, evaluation etc. A discourse is a sequence of sentences. Discourse has structure much like sentences do. Understanding discourse structure is extremely important for dialog system.

For example: The dialog may be

When does the bus to Bhubaneswar leave?

There is one at 10 a.m. and one at 1 p.m.

Give me two tickets for the earlier one, please.

The problems with discourse analysis may be non-sentential utterances, cross-sentential anaphora.

**Pragmatic Knowledge**

This level is concerned with the purposeful use of language in situations and utilizes context over and above the contents of the text for understanding. The goal is to explain how extra meaning is read into texts without actually being encoded in them. This requires much world knowledge including the understanding of intentions, plans and goals. Some NLP applications may utilize knowledge bases and inferencing modules. Pragmatic is the study of how more gets communicated than is said. Speech acts in the pragmatic processing is the illocutionary force, the communicative force of an utterance, resulting from the function associated with it. For example: Suppose the sentence is I will see you later.

Prediction: I predict that I will see you later.

Promise: I promise that I will see you later.

Warning: I warn you that I will see you later.

**EXAMPLES OF SOME NLP SYSTEMS**

In early 1950s, few NLP systems had been developed. Except the theoretical developments many practical systems were developed to demonstrate the effectiveness of particular principles. Weizenbaum’s ELIZA was built to replicate the conversation between a psychologist and a patient; simply by permuting the user input. Winograd’s SHRDLU simulated a robot that manipulated blocks on a table top. Also LUNAR was developed by Woods as an interface system to a database. In the late 1970’s, McKeown’s discourse planner TEXT and McDonald’s response generator MUMMBLE used theoretical predicates to produce declarative descriptions in the form of short texts, usually paragraphs. Some of the earliest NLP systems are described below.

# MCQs & answers

**1.** Natural language processing is field of\_\_\_\_\_\_\_\_\_\_\_\_\_.

(a) Computer science

(b) Artificial intelligence

(c) Linguistics

(d) All of the above

**2.** NLP is concerned with the interactions between computers and human (natural) languages.

(a) True

(b) False

**3.** One of the main challenge/s of NLP is \_\_\_\_\_\_\_\_\_\_\_\_\_.

(a) Handling ambiguity of sentences

(b) Handling tokenisation

(c) Handling POS-tagging

(d) All of the above

**4.** Choose from the following areas where NLP can be useful.

(a) Automatic text summarisation

(b) Automatic question–answering systems

(c) Information retrieval

(d) All of the above

**5.** The major tasks of NLP includes \_\_\_\_\_\_\_\_\_\_\_\_\_.

(a) Automatic summarisation

(b) Discourse analysis

(c) Machine translation

(d) All of the above

**6.** Machine translation

(a) Converts one human language to another human language

(b) Converts human language to machine language

(c) Converts any human language to English

(d) Converts machine language to human language

**7.** Optical Character Recognition uses NLP.

(a) True

(b) False

**8.** Information retrieval and Information Extraction are the two same thing.

(a) True

(b) False

**Answers**

**1. (d) 2. (a) 3. (a) 4. (d) 5. (d) 6. (a) 7. (a) 8. (b)**