**Assignment 5.1**

**1. List the various libraries and the functions related to the various steps of NLP.**

Ans)

Natural Language Processing(NLP), a field of AI, aims to understand the semantics and connotations of natural human languages. It focuses on extracting meaningful information from text and train data models based on the acquired insights. The primary NLP functions include text mining, text classification, text analysis, sentiment analysis, word sequencing, speech recognition & generation, machine translation, and dialog systems, to name a few.

1. Natural Language Toolkit (NLTK)

NLTK is one of the leading platforms for building Python programs that can work with human language data. It presents a practical introduction to programming for language processing. NLTK comes with a host of text processing libraries for sentence detection, tokenization, lemmatization, stemming, parsing, chunking, and POS tagging.

2. Gensim

Gensim is a Python library designed specifically for “topic modeling, document indexing, and similarity retrieval with large corpora.” All algorithms in Gensim are memory-independent, w.r.t., the corpus size, and hence, it can process input larger than RAM. With intuitive interfaces, Gensim allows for efficient multicore implementations of popular algorithms, including online Latent Semantic Analysis (LSA/LSI/SVD), Latent Dirichlet Allocation (LDA), Random Projections (RP), Hierarchical Dirichlet Process (HDP) or word2vec deep learning.

3. CoreNLP

Stanford CoreNLP comprises of an assortment of human language technology tools. It aims to make the application of linguistic analysis tools to a piece of text easy and efficient. With CoreNLP, you can extract all kinds of text properties (like named-entity recognition, part-of-speech tagging, etc.) in only a few lines of code.

4. spaCy

spaCy is an open-source NLP library in Python. It is designed explicitly for production usage – it lets you develop applications that process and understand huge volumes of text.

5. TextBlob

TextBlob is a Python (2 & 3) library designed for processing textual data. It focuses on providing access to common text-processing operations through familiar interfaces. TextBlob objects can be treated as Python strings that are trained in Natural Language Processing.

6. Pattern

Pattern is a text processing, web mining, natural language processing, machine learning, and network analysis tool for Python. It comes with a host of tools for data mining (Google, Twitter, Wikipedia API, a web crawler, and an HTML DOM parser), NLP (part-of-speech taggers, n-gram search, sentiment analysis, WordNet), ML (vector space model, clustering, SVM), and network analysis by graph centrality and visualization.

**2. List the various applications of NLP. Explain any two in detail.**

Ans)

**Email filters**

Email filters are one of the most basic and initial applications of NLP online. It started out with spam filters, uncovering certain words or phrases that signal a spam message. But filtering has upgraded, just like early adaptations of NLP. One of the more prevalent, newer applications of NLP is found in Gmail's email classification. The system recognizes if emails belong in one of three categories (primary, social, or promotions) based on their contents. For all Gmail users, this keeps your inbox to a manageable size with important, relevant emails you wish to review and respond to quickly.

**Smart assistants**

Smart assistants like Apple’s Siri and Amazon’s Alexa recognize patterns in speech thanks to voice recognition, then infer meaning and provide a useful response. We’ve become used to the fact that we can say “Hey Siri,” ask a question, and she understands what we said and responds with relevant answers based on context. And we’re getting used to seeing Siri or Alexa pop up throughout our home and daily life as we have conversations with them through items like the thermostat, light switches, car, and more. We now expect assistants like Alexa and Siri to understand contextual clues as they improve our lives and make certain activities easier like ordering items, and even appreciate when they respond humorously or answer questions about themselves. Our interactions will grow more personal as these assistants get to know more about us. As a New York Times article “Why We May Soon Be Living in Alexa’s World,” explained: “Something bigger is afoot. Alexa has the best shot of becoming the third great consumer computing platform of this decade.”

**Search results**

Search engines use NLP to surface relevant results based on similar search behaviors or user intent so the average person finds what they need without being a search-term wizard. For example, Google not only predicts what popular searches may apply to your query as you start typing, but it looks at the whole picture and recognizes what you’re trying to say rather than the exact search words. Someone could put a flight number in Google and get the flight status, type a ticker symbol and receive stock information, or a calculator might come up when inputting a math equation. These are some variations you may see when completing a search as NLP in search associates the ambiguous query to a relative entity and provides useful results.

**Predictive text**

Things like autocorrect, autocomplete, and predictive text are so commonplace on our smartphones that we take them for granted. Autocomplete and predictive text are similar to search engines in that they predict things to say based on what you type, finishing the word or suggesting a relevant one. And autocorrect will sometimes even change words so that the overall message makes more sense. They also learn from you. Predictive text will customize itself to your personal language quirks the longer you use it. This makes for fun experiments where individuals will share entire sentences made up entirely of predictive text on their phones. The results are surprisingly personal and enlightening; they’ve even been highlighted by several media outlets.

**Language translation**

One of the tell-tale signs of cheating on your Spanish homework is that grammatically, it’s a mess. Many languages don’t allow for straight translation and have different orders for sentence structure, which translation services used to overlook. But, they’ve come a long way. With NLP, online translators can translate languages more accurately and present grammatically-correct results. This is infinitely helpful when trying to communicate with someone in another language. Not only that, but when translating from another language to your own, tools now recognize the language based on inputted text and translate it.

**Digital phone calls**

We all hear “this call may be recorded for training purposes,” but rarely do we wonder what that entails. Turns out, these recordings may be used for training purposes, if a customer is aggrieved, but most of the time, they go into the database for an NLP system to learn from and improve in the future. Automated systems direct customer calls to a service representative or online chatbots, which respond to customer requests with helpful information. This is a NLP practice that many companies, including large telecommunications providers have put to use. NLP also enables computer-generated language close to the voice of a human. Phone calls to schedule appointments like an oil change or haircut can be automated, as evidenced by this video showing Google Assistant making a hair appointment.

**3. Describe the different stages of NLP.**

**Ans)**

**Morphological Analysis/ Lexical Analysis**

Morphological or Lexical Analysis deals with text at the individual word level. It looks for morphemes, the smallest unit of a word. For example, irrationally can be broken into ir (prefix), rational (root) and -ly (suffix). Lexical Analysis finds the relation between these morphemes and converts the word into its root form. A lexical analyzer also assigns the possible Part-Of-Speech (POS) to the word. It takes into consideration the dictionary of the language.

For example, the word “character” can be used as a noun or a verb.

**Syntax Analysis**

Syntax Analysis ensures that a given piece of text is correct structure. It tries to parse the sentence to check correct grammar at the sentence level. Given the possible POS generated from the previous step, a syntax analyzer assigns POS tags based on the sentence structure.

For example:

Correct Syntax: Sun rises in the east.

Incorrect Syntax: Rise in sun the east.

**Semantic Analysis**

Consider the sentence: “The apple ate a banana”. Although the sentence is syntactically correct, it doesn’t make sense because apples can’t eat. Semantic analysis looks for meaning in the given sentence. It also deals with combining words into phrases.

For example, “red apple” provides information regarding one object; hence we treat it as a single phrase. Similarly, we can group names referring to the same category, person, object or organisation. “Robert Hill” refers to the same person and not two separate names – “Robert” and “Hill”.

**Discourse**

Discourse deals with the effect of a previous sentence on the sentence in consideration. In the text, “Jack is a bright student. He spends most of the time in the library.” Here, discourse assigns “he” to refer to “Jack”.

**Pragmatics**

The final stage of NLP, Pragmatics interprets the given text using information from the previous steps. Given a sentence, “Turn off the lights” is an order or request to switch off the lights.