**Administrative**

|  |  |
| --- | --- |
| S/No | Items |
| 1 | Flask Attendance |
| 2 | Start recording |
| 3 |  |

**Concepts**

|  |  |
| --- | --- |
| S/No | Items |
|  | Coreference resolution is the task of finding all expressions that refer to the same entity in a text. It is an important step for a lot of higher level NLP tasks that involve natural language understanding such as document summarization, question answering, and information extraction. |
|  | OpenNLP – Java Based NLP toolkit  Pytorch NLP – Target at use of Deep Learning techniques for NLP tasks  Spacy / NLTK – Deals with basic tasks of NLP.  FastText – more of a specialist Library for efficient text classification and representation learning.  Gensim is a Python library for topic modelling, document indexing and similarity retrieval with large corpora and information retrieval (IR) community.All algorithms are **memory-independent** w.r.t. the corpus size (can process input larger than RAM, streamed, out-of-core), |
|  | **Parsing in NLP** is the process of determining the syntactic structure of a text by analyzing its constituent words based on an underlying grammar (of the language) |
|  | **Example of trained models in NLTK**  [**VADER**](https://github.com/cjhutto/vaderSentiment) (**Valence Aware Dictionary and sEntiment Reasoner**) is a lexicon and rule-based sentiment analysis tool that isspecifically attuned to sentiments expressed in social media. |
|  |  |
|  | **WordNet Database / WordNetLemmistizer**  WordNet® is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations.  WordNet superficially resembles a thesaurus, in that it groups words together based on their meanings. However, there are some important distinctions. First, WordNet interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than meaning similarity.  **More details at:** [**https://wordnet.princeton.edu/**](https://wordnet.princeton.edu/) |
|  | **Penn Treebank**  In [linguistics](https://en.wikipedia.org/wiki/Linguistics), a treebank is a parsed [text corpus](https://en.wikipedia.org/wiki/Text_corpus) that [annotates](https://en.wikipedia.org/wiki/Annotated) [syntactic](https://en.wikipedia.org/wiki/Syntactic) or [semantic](https://en.wikipedia.org/wiki/Semantic) [sentence](https://en.wikipedia.org/wiki/Sentence_(linguistics)) linguistic structure such as Part of Speech tagging.  The treebank provides insights into the “what’s” of a writing text through tagging, syntactic descriptions of sentences often come in the form of tree structures  The Penn Treebank Project annotates naturally-occurring text for linguistic structure. The Penn Treeban produced approximately 7 million words of part-of-speech tagged text, 3 million words of skeletally parsed text, over 2 million words of text (but based on WSJ).  Ref: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.9.8216&rep=rep1&type=pdf> |
|  |  |

**Concepts / Further Explanation**

|  |  |
| --- | --- |
| S/No | Items |
|  |  |

**Supplementary Resources**

|  |  |
| --- | --- |
| S/No | Resource details |
|  |  |