

Activity 1

Natural language processing (NLP) is an interdisciplinary subfield of computer science and linguistics. It is primarily concerned with giving computers the ability to support and manipulate human language. It involves processing natural language datasets, such as text corpora or speech corpora, using either rule-based or probabilistic (i.e. statistical and, most recently, neural network-based) machine learning approaches. The goal is a computer capable of "understanding" the contents of documents, including the contextual nuances of the language within them. The technology can then accurately extract information and insights contained in the documents as well as categorize and organize the documents themselves.

1. RE to find all **digits** in above text: _____ Count #of digits: _____
2. RE to find all **'it'** in the above text: _____
3. Words containing **'it'** only in the text, it's count: _____ and RE _____
4. Words containing **'the'** only from the above text. RE: _____
5. RE to find all words that doesn't contains any Letter or Digit _____ in the text.