Unit I

S.NO	Question
1	Explain how the Stop word removal and Rare word removal are performed using NLTK with examples.
2	Compute the minimum edit distance between intention and execution using the minimum edit distance algorithm.
3	How Text wrangling and Cleansing are performed by using NLTK.
4	Explain in detail about basic regular expression patterns
5	Write regular expressions for the following languages. 1. the set of all alphabetic strings; 2. the set of all lower case alphabetic strings ending in a b; 3. the set of all strings from the alphabet a,b such that each a is immediately preceded by and immediately followed by a b 4. The set of all strings with two consecutive repeated words.
6	Explain Byte-Pair Encoding for Tokenization with an Example
7	What are the Unix Tools used for Crude Tokenization and Normalization?
8	Write a RE to find cases of the English article “the”.
9	Explain how Tokenization, Stemming, Lemmatization will be done using NLTK.