

Worksheet

Name:
Roll No: Class: Section:
Date:

NATURAL LANGUAGE PROCESSING

Chapter-6

1. Fill in the blanks.

- is the number of documents in which the word occurs irrespective of how many times it has occurred in those documents.
- NLTK stands for
- TFIDF stands for
- is the process of dividing the sentences further into tokens.
- is a process of Sentence Boundary Detection which reduces the corpus into a sentence.
- In summarisation, the selected text, phrases, sentences or sections are picked up from the scattered resources and joined appropriately to form a concise summary.

2. Differentiate following:

- Human Language and Computer Language

.....
.....
.....

- Script-bot and Smart-bot

.....
.....
.....

- Data Acquisition and Data Exploration

.....
.....
.....