



National University
Of Computer and Emerging Sciences

Name:

Dawood Sarfraz

Roll no:

20P-0153

Section:

BSCS-7A

Course:

Natural Language Processing

Lab:

02

Submitted to:

Dr. Omer Usman Khan

FAST National University of Computer and Emerging Sciences

Table - 1:

	Text1	Text2	Tex3
10 frequently occurring Bigrams	Sperm Whale; Moby Dick; White Whale; old man; Captain Ahab; sperm whale; Right Whale; Captain Peleg; New Bedford; Cape Horn;	Colonel Brandon; Sir John; Lady Middleton; Miss Dashwood; every thing; thousand pounds; dare say; Miss Steeles; said Elinor; Miss Steele;	said unto; pray thee; thou shalt; thou hast; thy seed; years old; spake unto; thou art; LORD God; every living;
5 frequently occurring Trigrams	[('AFTER', 'EXCHANGING', 'HAILS'), ('Anacharsis', 'Cloutz', 'deputation'), ('CAULKING', 'ITS', 'SEAMS'), ('ELIZABETH', 'OAKES', 'SMITH'), ('Et', 'tu', 'Brute')]	[('Austen', '1811', ' '), ('Jane', 'Austen', '1811'), ('200', 'L', 'per'), ('Drury', 'Lane', 'lobby'), ('L', 'per', 'annum')]	[('olive', 'leaf', 'pluckt'), ('sewed', 'fig', 'leaves'), ('yield', 'royal', 'dainties'), ('Fifteen', 'cubits', 'upward'), ('leaf', 'pluckt', 'o')]
Number of words with length > 16	['cannibalistically', 'characteristically', 'circumnavigations', 'comprehensiveness', 'indispensableness', 'preternaturalness', 'subterraneousness', 'superstitiousness', 'uncomfortableness', 'uncompromisedness', 'uninterpenetratingly']	['companionableness', 'disinterestedness', 'disqualifications']	[] No word found
Number of words with Frequency > 500	['Ahab', 'But', 'I', 'The', 'a', 'all', 'an', 'and', 'are', 'as', 'at', 'be', 'but', 'by', 'for', 'from', 'had', 'have', 'he', 'him', 'his', 'in', 'into', 'is', 'it', 'like', 'man', 'me', 'more', 'my', 'not', 'now', 'of', 'on', 'one', 'or', 'out', 's', 'ship', 'so', 'some', 'that', 'the', 'their', 'then', 'there', 'they', 'this', 'to', 'up', 'upon', 'was', 'were', 'whale', 'when', 'which', 'with', 'you']	['I', 'The', 'a', 'all', 'an', 'and', 'are', 'as', 'at', 'be', 'but', 'by', 'for', 'from', 'had', 'have', 'he', 'him', 'his', 'in', 'into', 'is', 'it', 'like', 'man', 'me', 'more', 'my', 'not', 'now', 'of', 'on', 'one', 'or', 'out', 's', 'so', 'some', 'that', 'the', 'their', 'then', 'there', 'they', 'this', 'to', 'up', 'upon', 'was', 'were', 'when', 'which', 'with', 'you']	['I', 'The', 'a', 'all', 'an', 'and', 'are', 'as', 'at', 'be', 'but', 'by', 'for', 'from', 'had', 'have', 'he', 'him', 'his', 'in', 'into', 'is', 'it', 'like', 'man', 'me', 'more', 'my', 'not', 'now', 'of', 'on', 'one', 'or', 'out', 's', 'so', 'some', 'that', 'the', 'their', 'then', 'there', 'they', 'this', 'to', 'up', 'upon', 'was', 'were', 'when', 'which', 'with', 'you']

Number of words ending in “ed”	2196	902	281
---------------------------------------	------	-----	-----

Table - 2:

Number of Words In Corpus	30 word generated sentence
1628	computer vision , and generate human language . The future trajectory of NLP . The future trajectory of NLP . The future trajectory of NLP . The future trajectory of
1628	NLP . The future trajectory of NLP . The future trajectory of NLP . The future trajectory of NLP . The future trajectory of NLP . The future trajectory of

```
In [93]: generate_model(cfd, 'computer', 30)
```

```
computer vision , and generate human language . The future trajectory of NLP . The future trajectory of NLP . The future trajectory of NLP . The future trajectory of
```

```
In [94]: len(wordlist)
```

```
Out[94]: 1628
```

```
In [91]: generate_model(cfd, 'NLP', 30)
```

```
NLP . The future trajectory of NLP . The future trajectory of NLP . The future trajectory of NLP . The future trajectory of NLP . The future trajectory of
```

```
In [92]: len(wordlist)
```

```
Out[92]: 1628
```