**Charotar University of Science and Technology [CHARUSAT]**

**Chandubhai S. Patel Institute of Technology [CSPIT]**

**U & P U. Patel Department of Computer Engineering**

**Lab Manual**

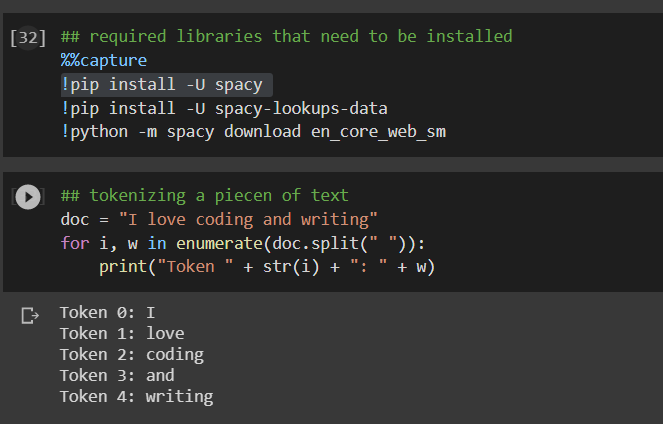
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subject code | : | CE-350 | Semester | : | 6 | Academic Year | : | 2021 |
| Subject name | : | Data Warehouse and Data Mining | | | | | | |

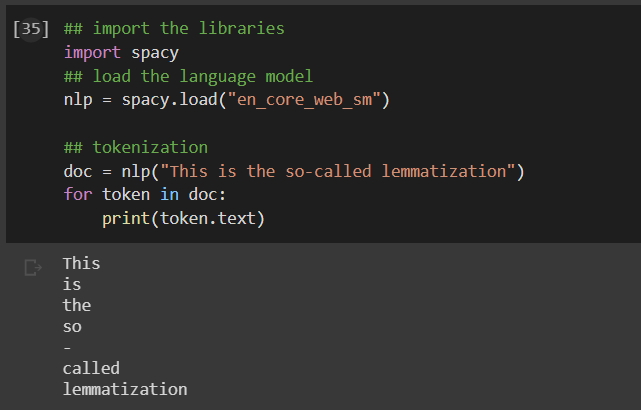
**Practical – 9,10,11**

**Aim:**

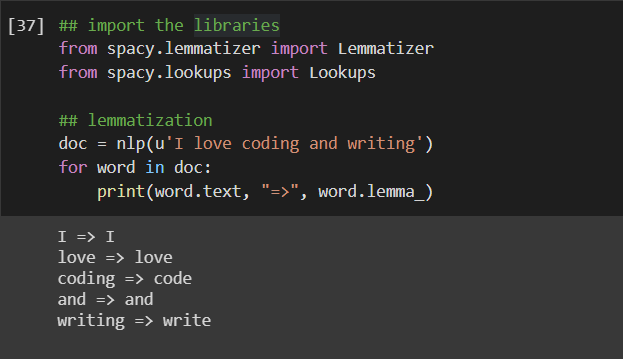
Perform Tokenization, Lemmatization, Stemming, and Sentence Segmentation in Google colab

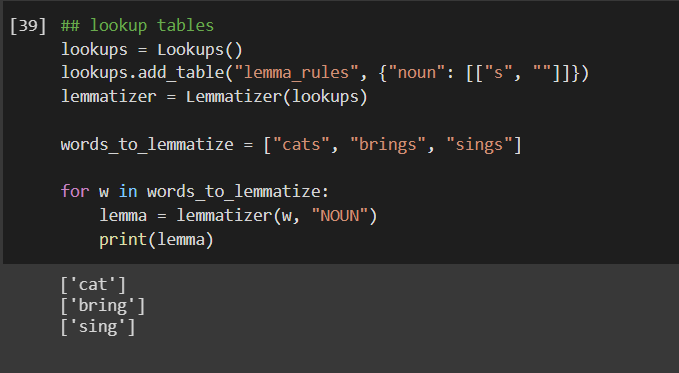
**Tokenization:-**



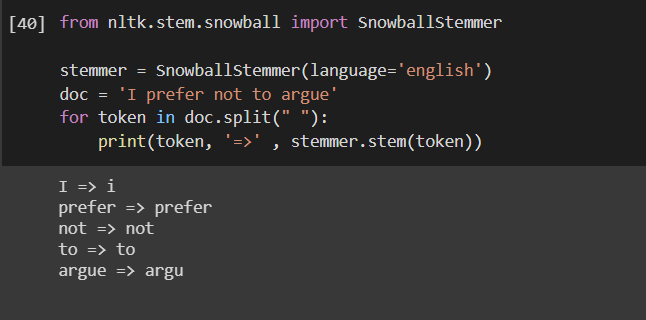


**Lemmatization**

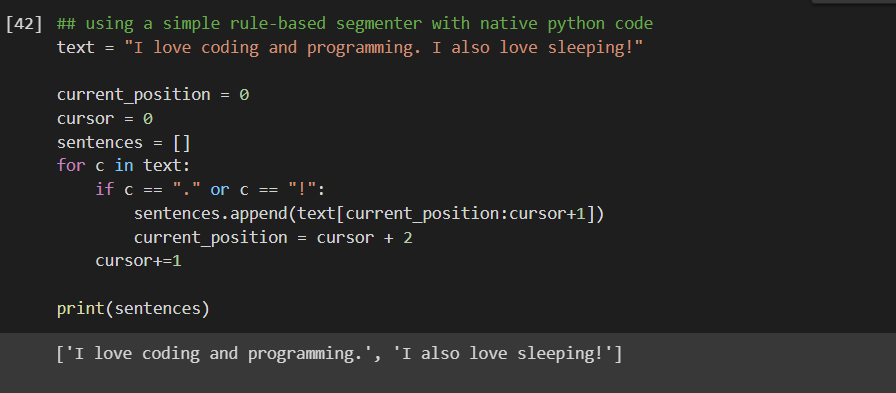


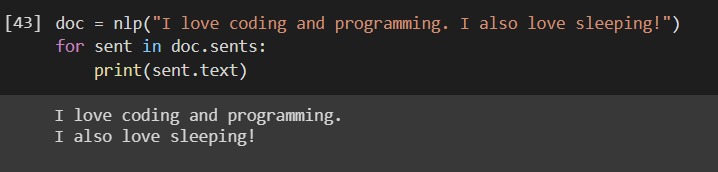


**Stemming**



**Sentence Segmentation**





|  |  |  |  |
| --- | --- | --- | --- |
| **Prepared By:** | Rajiv Kumar Gupta (18CE137) | **Date:** | 04-30-2021 |