İstanbul Bilgi University

Department of Computer Engineering

Spring 2018-2019

CMPE 346 : Natural Language Processing

Term Project - Phase III

23rd April, 2019 23:59

1. The term project is composed of several phases.

2. Second phase is to investigate the language model of the collected data.

3. For each phase of the term project, you need to make a submission before the deadline.

4. After the submission of the regarding step, you need to be present during the determined lab section and make a demonstration to explain your work and to show your progress.

5. The choices of the programming language and the development environment are up to you. However, it is highly encouraged to use Pyhton and NLTK package.

6. NLTK is quite popular and widely used open source package for text analysis and natural language processing purposes.

7. You can refer to the sample programs that we have done during the lab sessions.

8. The official website of the NLTK[1] contains practical tutorials, samples and documentation which might be really beneficial for the project.

9. Please check announcements and file uploads regularly for the upcoming steps of the project.

---

[1]http://www.nltk.org/

Steps of Phase III

1. You are expected to create a function that constructs a lexicon that maps words of the English language to lexical categories.

2. You are expected to create a function which constructs a grammar that defines the structure of a sentence.

3. You are expected to create a parser function, which applies the CYK algorithm, takes a grammar that is produced from a textual data as parameter.