Python_Lesson7: Python Programming

Lesson Overview:
In this lesson we will focus on text processing like unigram, bigram, trigram, tokenization, pos tagging, lemmatization, normalization, entity extraction, language model. Learning these features will help us for more meaningful project as document classification, spelling corrector, document summarization, etc.

Use Case Description:
In this use case, we will learn how to correct the mistyped of words in a sentence.
Spelling corrector is about using some of the NLP features we learned during the class, then correcting a mistyped word. Thus, students can see the right application of these features in a project.

Programming elements:
Basic NLP techniques like unigram, bigram, trigram, tokenization, pos tagging, lemmatization, normalization, entity extraction, language model

Source Code:
https://umkc.box.com/s/8vygyn9iqj8ut6k8vn434jmpfoldde20

In class programming:
In class, we further work on the tokenization, pos-tagging, entity extraction, bigram and trigram.
For all the exercises import the right module from NLTK. You need to go through the slides to find them.

1. Change the classifier in the given code to
   a. **SVM** and see how accuracy changes
   b. change the tfidf vectorizer to use bigram and see how the accuracy changes
   TfidfVectorizer(ngram_range=(1,2))
   c. Set argument stop_words='english' and see how accuracy changes

2. Extract the following web URL text using BeautifulSoup
https://en.wikipedia.org/wiki/Google
3. Save it in **input.txt**

4. Apply the following on the **"input.txt"** and show output:
   a. Tokenization
   b. POS
   c. Stemming
   d. Lemmatization
   e. Trigram
   f. Named Entity Recognition

Evaluation Criteria:
1. Completeness of Features
2. Code Quality (https://en.wikipedia.org/wiki/Best_coding_practices)
3. Time
4. Feedback Submission

Note: *Cheating, plagiarism, disruptive behavior and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL:*
https://catalog.umkc.edu/special-notices/academic-honesty/