START OF QUIZ
Student ID:
30542179,Yang,Yimei

Question 1

Topic: Lecture 1
Source: Lecture 1

In class, we talked about how .isdigit() is insufficient for determining whether we can convert a string to a float. Write a short function "isfloat" that determines whether a provided string is a valid floating point number. (2)

Question 2

Topic: Lecture 3
Source: Lecture 3

Imagine that we have a parallel corpus (ie, a corpus containing sentences in two languages), and we want to extract a bilingual lexicon. What are some simple steps we could do to identify words that could be translations of each other? (2)

Question 3

Topic: Lecture 3
Source: Lecture 3

Although lexicons are often good starting points, they are often less capable than ML methods. What are some reasons (at least 2) that lexicons are insufficient for state-of-the-art training. Briefly explain. (2)

Question 4

Topic: Lecture 2
Source: Lecture 2

How does Zipf's law relate to Hapax Legomena? (1)

Question 5

Topic: Lecture 2
Source: Lecture 2

As we expand the size of our corpus, we increase the number of Hapaxes. Do you think this is also true of stopwords? Briefly explain. (1)

Question 6

Topic: Lecture 1
Source: Lecture 1

How would you quickly identify the words in a sentence? (1)

Question 7

Topic: Lecture 4
Source: Lecture 4

We discussed two alternative methods for noise reduction: removing all words above a certain frequency, or only removing those from a curated lexicon. Name an advantage to both. (1)

Question 8

Topic: Lecture 4
Source: Lecture 4

Do you think that children's (age 3-5) picture books would have a higher or lower ratio of adjectives than university literature? Briefly explain your logic. (1)

Question 9

Topic: Coding
Source: Coding

Grimm's law is a linguistic phenomenon that describes how sounds in language (mostly related to the Germanic languages like English, Dutch, German, Norwegian, Icelandic, etc.) changed over time (specifically from some progenitor thousands of years old - Germanic languages observed the change, while Romance languages did not). For example, the /p/ sound in Latin evolved into the /f/ sound in English across many words - compare "piscus" with "fish"; "pater" with "father"; "pedus" with "foot" (there are a couple other changes in there, too - see if you can spot them!). If you had a time machine, and could bring a computational toolkit to help Jacob Grimm formulate his law, what would you need, in terms of lexicons, keeping in mind that the /p/ -> /f/ change is only one of a handful of sound changes, and that the changes occured over dozens of languages? Explain (with pseudocode, if necessary), how you would start to identify trends in the data (assuming that your computer still works in the 19th century)? (3)

END OF QUIZ