START OF QUIZ
Student ID: 47881305,Hrabowsky,Zenon

Question 1

Topic: Lecture 1
Source: Lecture 1

Vowels are often used as a proxy for syllables in words (it's not a perfect correspondence, but it's not bad). Write a function that counts the vowels in a word, without using a loop, using only the tools we went over in Lecture 1 (list comprehension counts as a loop). (2)

Question 2

Topic: Lecture 3
Source: Lecture 3

How does ”get” differ from a default dictionary (2 ways)? (1)

Question 3

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 4

Topic: Lecture 4
Source: Lecture 4

In French, negation is often indicated by ”ne ... pas” (ie, ”je ne parle pas” - ”I am not speaking”; ”tu ne conduis pas” - ”You are not driving”, etc.). However, in speech, one of the two is often dropped: ”je ne parle.” or ”tu conduis pas.”. Using this information, how would you determine whether a corpus was composed of written or spoken French? You don’t need to write the code, but explain the logic that you would use to come to this conclusion. (2)

Question 5

Topic: Lecture 2
Source: Lecture 2

If we have a new corpus, how might we automatically determine (without ML): A. The language it's written in. B. Whether it is annotated C. If it is multilingual D. genre? Briefly explain your reasoning. (2)

Question 6

Topic: Lecture 3
Source: Lecture 3

What properties of dictionaries make them an efficient choice for nesting complex lexicons.
(1)

Question 7

Topic: Lecture 1
Source: Lecture 1

What are two ways to check if a word is all capital letters (neither should require more than one function call)? (1)

Question 8

Topic: Lecture 2
Source: Lecture 2

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

Question 9

Topic: Coding
Source: Lecture 2

Imagine we have a large corpus in an unknown language. We don't have any ML tools to analyze the data. How might we determine the stopwords in our corpus? How might we test our theory of stopwords? (I'll make it easy on you - the tokens are space separated, and we have some way of separating sentences.) (3)

END OF QUIZ