START OF QUIZ Student ID: 34447581,Ong,Claudia

Topic: Lecture 2 Source: Lecture 2

Why is it important to know when a corpus was constructed, and who constructed it? (1)

Topic: Lecture 3 Source: Lecture 3

Why do we not care about the extra space required to create a reverse index? (2 reasons) (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)

Topic: Lecture 4 Source: Lecture 4

Attributive adverbs are a type of adverb that provides "flavour" to speech verbs (example: "she said quickly"; "he spoke loudly"). They are often frowned upon in formal writing, because they can be replaced with other verbs: "blurted" or "shouted", in the example. Write a quick function that finds them in the Brown corpus, and reports how many sentences in 1000 have them. (2)

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)

Topic: Lecture 3 Source: Lecture 3

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

(1)

Topic: Lecture 4 Source: Lecture 4

Why does type-to-token ratio decrease as the size of the corpus increases? What does this suggest about long documents? (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)

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