

**START OF QUIZ**

**Student ID:**

**36304153,Kang,David**

## Question 1

Topic: Lecture 1

Source: Lecture 1

How would you convert a string into a list of characters? (1)

## Question 2

Topic: Lecture 4

Source: Lecture 4

In class, we removed stopwords by using a lexicon. Can you think of another way that we could remove all closed class words? (1)

### Question 3

Topic: Lecture 1

Source: Lecture 1

What method would you use to check if a string contains only numeric digits (including decimals) without using any additional libraries? (1)

## Question 4

Topic: Lecture 2

Source: Lecture 2

How does Zipf's law help explain the distribution of word frequencies in a corpus? What impacts does that have on our algorithms? (1)

## Question 5

Topic: Lecture 4

Source: Lecture 4

Would you expect a higher or lower frequency of passive voice constructions in legal documents compared to casual conversation? Briefly explain your reasoning. (Remember that passive voice is a structure like "the tree was cut down", inverting the subject and object). (1)

## Question 6

Topic: Lecture 3

Source: Lecture 3

Lexicons are useful for initial text analysis but often lack the adaptability needed for advanced NLP tasks. Why is this the case? Provide at least 2 reasons with brief explanations.  
(2)

## Question 7

Topic: Lecture 3

Source: Lecture 3

Imagine you have a large text corpus in English and Spanish and want to automatically align sentences for machine translation. What are some straightforward methods you could use to identify sentence pairs that are likely translations of each other? (2)



## Question 8

Topic: Lecture 2

Source: Lecture 2

Is it possible for a corpus of a low-resource language to follow Zipf's law? What factors might influence the degree to which the law applies in such languages? (2)

## Question 9

Topic: Long

Source: Lecture 1

Write a function that validates if a string matches a phone number format, such as (123) 456-7890. What types of invalid inputs should the function check for? Are there edge cases we would be willing to accept? How would we handle those? Write 3 test cases - 2 that should pass, and one that should fail. (3)

**END OF QUIZ**