MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

P.O. Box 972-60200 – Meru-Kenya
Tel: +254(0) 799 529 958, +254(0) 799 529 959, + 254 (0) 712 524 293,
Website: info@must.ac.ke Email: info@must.ac.ke

University Examinations 2024/2025

FOURTH YEAR FIRST SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF DATA SCIENCE, BACHELOR OF COMPUTER SCIENCE

CDS 3400: NATURAL LANGUAGE PROCESSING

DATE: DECEMBER 2024 **TIME: 2 HOURS**

INSTRUCTIONS: *Answer question **ONE** (Compulsory) and any other **TWO** questions*

QUESTION ONE (30 MARKS)

a) Define sentiment analysis in the context of natural language processing. Explain its applications and significance. (5 marks)

b) Discuss the differences between lexicon-based and machine learning approaches in sentiment analysis. Provide examples of each approach. (5 marks)

c) Implement a basic lexicon-based sentiment analyzer in Python that can classify a given sentence as positive, negative, or neutral. (5 marks)

d) Explain the concept of overfitting in the context of sentiment analysis with machine learning models. Discuss techniques to prevent overfitting. (5 marks)

e) Share your thoughts on the potential challenges and biases in sentiment analysis when applied to social media data. (5 marks)

f) Calculate the perplexity of a given sentence using a language model. Show your calculations. (5 marks)

QUESTION TWO (20 MARKS)

a) Define dependency parsing and explain the role of dependency trees in representing sentence structure. (5 marks)

b) Describe the transition-based parsing approach in dependency parsing. Provide an example of how a transition-based parser works. (5 marks)

c) Implement a basic dependency parser in Python that can parse a given sentence and display its dependency tree structure. (5 marks)

d) Discuss the advantages and disadvantages of dependency parsing compared to constituency parsing. (5 marks)

QUESTION THREE (20 MARKS)

a) Define Semantic Role Labeling (SRL) and explain its importance in natural language understanding. (5 marks)

b) Describe the concepts of Frame Net and Prop Bank in the context of SRL. Explain how they are used in annotating sentences with semantic roles. (5 marks)

c) Implement a Python code snippet to perform semantic role labeling for a provided sentence. (5 marks)

d) Discuss the challenges of automatic Semantic Role Labeling for languages with complex syntactic structures. Provide examples if necessary. (5 marks)

QUESTION FOUR (20 MARKS)

a) Explain the differences between rule-based and statistical approaches to machine translation. Provide examples of languages that might benefit from each approach. (5 marks)

b) Describe the concept of phrase-based machine translation and how it differs from word-based machine translation. (5 marks)

c) Implement a simple machine translation system in Python for a given language pair. (5 marks)

d) Discuss the challenges of machine translation for languages with highly inflected morphology. Provide examples if necessary. (5 marks)

QUESTION FIVE (20 MARKS)

a) Define dialogue systems and discuss their role in human-computer interaction. (5 marks)

b) Describe the difference between task-oriented and conversational dialogue systems. Provide examples of applications for each type. (5 marks)

c) Implement a basic task-oriented dialogue system in Python for a provided task. (5 marks)

d) Discuss the challenges of building dialogue systems that can handle natural language variations, user interruptions, and multiple intents in a conversation. (5 marks)