



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 2023/2024

FOURTH YEAR SECOND SEMESTER EXAMINATION FOR THE DEGREE OF
BACHELOR OF COMPUTER TECHNOLOGY

CDS 3400: NATURAL LANGUAGE PROCESSING

DATE: APRIL 2024

TIME: 2 HOURS

INSTRUCTIONS: Answer question *one* and any other *two* questions

QUESTION ONE (30 MARKS)

- a) Define Natural Language Processing (NLP) and explain its importance in modern computing. (5 Marks)
- b) Describe the tokenization process in NLP. Provide an example to illustrate your explanation. (5 Marks)
- c) Explain the concept of stemming in NLP. How does it differ from lemmatization? (5 Marks)
- d) Discuss the challenges faced in Named Entity Recognition (NER) and propose potential solutions to overcome these challenges. (5 Marks)
- e) What is the difference between syntactic and semantic parsing in NLP? Provide examples to distinguish between the two. (5 Marks)



MUST is ISO 9001:2015 and



ISO/IEC 27001:2013 CERTIFIED

- f) Define Word Embeddings and explain how they are useful in NLP tasks such as sentiment analysis and text classification. (5 Marks)

QUESTION TWO (20 MARKS)

- a) Explain the importance of sentiment analysis in the context of social media analytics. Provide examples of how sentiment analysis can be used to extract insights from social media data. (5 Marks)
- b) Describe two common approaches to sentiment analysis: lexicon-based and machine learning-based. Compare their strengths and weaknesses. (5 Marks)
- c) Discuss the challenges faced in sentiment analysis when dealing with sarcasm and irony in text. Propose strategies to address these challenges. (5 Marks)
- d) Provide a step-by-step explanation of how you would implement sentiment analysis using a machine learning approach. Include details on data preprocessing, feature extraction, model training, and evaluation. (5 Marks)

QUESTION THREE (20 MARKS)

- a) Define Named Entity Recognition (NER) and explain its role in information extraction from unstructured text data. (5 Marks)
- b) Discuss the types of named entities that can be recognized using NER systems. Provide examples of each type. (5 Marks)
- c) Explain the difference between rule-based and statistical approaches to NER. Discuss the advantages and disadvantages of each approach. (5 Marks)
- d) Describe a real-world application of NER in a specific domain (e.g., biomedical, finance, etc.). Discuss how NER enhances information retrieval or analysis in this domain. (5 Marks)

QUESTION FOUR (20 MARKS)

- a) Define text summarization and explain its importance in handling large volumes of textual data. Provide examples of scenarios where text summarization is beneficial. (5 Marks)



- b) Compare and contrast extractive and abstractive text summarization approaches. Discuss the trade-offs between these two methods. (5 Marks)
- c) Discuss the challenges in evaluating the quality of text summarization. Describe commonly used evaluation metrics and their limitations. (5 Marks)
- d) Explain how deep learning models such as Recurrent Neural Networks (RNNs) and Transformer-based architectures have improved the performance of abstractive text summarization. Provide examples of state-of-the-art models in this area. (5 Marks)

QUESTION FIVE (20 MARKS)

- a) Discuss the challenges in machine translation, including issues related to syntax, semantics, and cultural nuances. (5 Marks)
- b) Explain the difference between rule-based and statistical machine translation approaches. Provide examples of each approach. (5 Marks)
- c) Describe the architecture of a neural machine translation (NMT) model. Discuss how NMT addresses some of the limitations of traditional statistical approaches. (5 Marks)
- d) Compare the performance of NMT models with traditional statistical machine translation systems. Discuss the advantages and limitations of each approach in different language pairs and domains. (5 Marks)

