



MACHAKOS UNIVERSITY

University Examinations for 2022/2023

SCHOOL OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF COMPUTING AND INFORMATION TECHNOLOGY

FOURTH YEAR SECOND SEMSTER EXAMINATION FOR

BACHELOR OF SCIENCE COMPUTER SCIENCE

SCO 409: NATURAL LANGUAGE PROCESSING

DATE:

TIME:

INSTRUCTIONS:

This paper consists of FIVE questions

*Answer **question one** and other **two questions** in this paper*

QUESTION ONE (COMPULSORY) (30 MARKS)

- a) Natural language processing is difficult due to ambiguities and uncertainties. Distinguish between Anaphoric and Pragmatic Ambiguity (4 marks)
- b) In business application, discuss TWO tasks involved in Natural language understanding to recognize the customer's problem in both spoken and written language. (4 marks)
- c) Define a function `hedge(text)` that processes a text and produces a new version with the word 'like' between every third word. (4 marks)
- d) Discuss TWO Properties of Stochastic POST Tagging (4 marks)
- e) Consider the following Python expression: `len(set(text4))`. State the purpose of this expression. Describe the two steps involved in performing this computation. (4 marks)
- f) Write a function that finds the 50 most frequently occurring words of a text that are not stop word (5 marks)
- g) Explain FIVE applications of Natural Language Processing (5 marks)

QUESTION TWO (20 MARKS)

- a) Briefly discuss how each of the following NLP techniques can be used in extractive summarization.
 - i. morphological processing; (2 marks)
 - ii. syntactic parsing; (2 marks)
 - iii. lexical semantics (2 marks)
 - iv. discourse parsing (2 marks)
- b) Explain the THREE steps involved in Pseudo Feedback (4 marks)
- c) The general problem of anaphora resolution includes finding antecedents for definite expressions, such as the jacket having his clothes as an antecedent.
 - i. Outline how one might exploit semantic relations from WordNet as part of an extension of the classifier approach to deal with definite expressions. (4 marks)
 - ii. What limitations might WordNet have for this task? (4 marks)

QUESTION THREE (20 MARKS)

- a) Describe the TWO most common types of Treebanks in linguistics. (4 marks)
- b) A semantic system brings entities, concepts, relations and predicates together to provide more context to language so machines can understand text data with more accuracy. Explain THREE elements that help to support language understanding (6 marks)
- c) Define a string `s = 'colorless'`. Write a Python statement that changes this to "colourless" using only the slice and concatenation operations. (5 marks)
- d) State and explain the two-stage architecture of Rule-based POS tagging (5 marks)

QUESTION FOUR (20 MARKS)

- a) Discuss FIVE roles of main roles of the parser (5 marks)
- b) Explain the FOUR components of Context Free Grammar (4 marks)
- c) Highlight FOUR advantages of Transformation-based Learning (TBL) (4 marks)
- d) Write regular expressions to match the following classes of strings: (7 marks)
 - i. A single determiner (assume that a, an, and the are the only determiners)
 - ii. An arithmetic expression using integers, addition, and multiplication, such as $2*3+8$

QUESTION FIVE (20 MARKS)

- a) Describe TWO steps involved in lexical semantics (4 marks)
- b) Explain THREE difficulties experienced in Word Sense Disambiguation (WSD) (6 marks)
- c) Discuss THREE design features of Information retrieval (IR) systems (6 marks)
- d) Write a python code to define a parser for parsing the grammar. (4 marks)