

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Register Number** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

***SRM Nagar, Kattankulathur – 603203, Chengalpattu District, Tamil Nadu***

**College of Engineering and Technology**

**School of Computing**

**Department of Data Science and Business Systems**

|  |  |  |  |
| --- | --- | --- | --- |
| **Test** | **:** CLA-T2 | **Date** | **:** 05.04.2024 |
| **Course Code & Title** | **:** NATURAL LANGUAGE PROCESSING | **Duration** | **:** 2 HOURS |
| **Year & Sem** | **:** III Year / V Sem | **Max Marks** | **:** 50 |

**Course Articulation Matrix:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S.No.** | **Course Outcome** | **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** |
| 1 | CO1 | 3 | 3 | 2 | - | - | - | - | - | - | - | - | - |
| 2 | CO2 | 3 | 3 | 2 | - | - | - | - | - | - | - | - | - |
| 3 | CO3 | 3 | 3 | 2 | - | - | - | - | - | - | - | - | - |
| 4 | CO4 | 3 | - | - | 3 | 3 | - | - | - | - | - | - | - |
| 5 | CO5 | - | - | 2 | 3 | 3 | - | - | - | - | - | - | - |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Part – A (10 x 1 = 10 Marks)**  **Instructions: 1) Answer ALL questions. 2) Encircle the correct answers. 3) MCQs sheet will be collected after 20 minutes.** | | | | | | |
| Q. No | Question | Marks | BL | CO | PO | PI Code |
| 1 | Identify the ambiguity of statement “**Visiting relatives can be annoying**“   1. Lexical ambiguity 2. Syntactic ambiguity 3. Semantic ambiguity 4. Discourse ambiguity | 1 | L1 | 2 | 1 | 1.6.1 |
| 2 | A DFA is a tuple A = (Q, ∑, δ, qo, F) ,what does δ indicates?   1. Finite set of state 2. A finite set of input symbols 3. Transition function 4. A set of final states | 1 | L1 | 2 | 1 | 1.6.1 |
| 3 | Context –free grammars also known as …..........   1. Meaning structure grammars 2. Character structure grammars 3. Shape structure grammars 4. Phrase structure grammars | 1 | L1 | 2 | 1 | 1.6.1 |
| 4 | What is the number of trigrams in a normalized sentence of length n words?   1. N 2. N-1 3. N-2 4. N-3 | 1 | L1 | 2 | 1 | 1.6.1 |
| 5 | In the pair "doctor-profession," what kind of lexical relation exists between the words?   1. Hyponymy 2. Synonymy 3. Meronymy 4. Antonymy | 1 | L1 | 2 | 1 | 1.6.1 |
| 6 | Which word embedding technique uses a neural network to learn word representations?   1. Latent Semantic Analysis 2. Word2Vec 3. Term Frequency-Inverse Document Frequency 4. Bag of Words | 1 | L1 | 2 | 1 | 1.6.1 |
| 7 | The process of removing words like “and”, “is”, “a”, “an”, “the” from a sentence is called as   1. Stemming 2. Lemmatization 3. Stop word Removal 4. POS | 1 | L1 | 2 | 1 | 1.6.1 |
| 8 | What is a "head word" in the context of syntax and parsing?   1. The main verb of a sentence 2. The first word in a sentence 3. A type of punctuation mark 4. The longest word in a sentence | 2 | L1 | 2 | 1 | 1.6.1 |
| 9 | In CFG, Each rule has a \_\_\_\_\_\_\_\_\_\_side   1. Right hand only 2. Left hand only 3. Right hand and left hand 4. Sub hand | 1 | L1 | 2 | 1 | 1.6.1 |
| 10 | Which technique is used to improve the accuracy of search results in semantic search?   1. Regular expressions 2. Word embeddings 3. Genetic algorithms 4. Markov chains | 1 | L1 | 3 | 1 | 1.6.1 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Register Number** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |



**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

***SRM Nagar, Kattankulathur – 603203, Chengalpattu District, Tamil Nadu***

**College of Engineering and Technology**

**School of Computing**

**Department of Data Science and Business Systems**

|  |  |  |  |
| --- | --- | --- | --- |
| **Test** | **:** CLA-T2 | **Date** | **:** 05.04.2024 |
| **Course Code & Title** | **:** NATURAL LANGUAGE PROCESSING | **Duration** | **:** 100 Minutes (2 Periods) |
| **Year & Sem** | **:** III Year / V Sem | **Max Marks** | **:** 50 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Part – B (4 x 10 = 40 Marks)**  **Instructions: Answer ANY four** | | | | | | |
| **Q. No** | **Question** | **Marks** | **BL** | **CO** | **PO** | **PI Code** |
| 11 | 1. Derive a top down parse tree for the given sentence:   **“The angry bear chased the frightened little squirrel”**  Use the following grammar rules to create the parse tree:  S-> NP VP Det -> the  NP->Det Noun Adj-> little | angry | frightened  VP-> V NP N-> Squirrel | bear  Nom-> Adj Noun | N V -> Chased     1. Perform parsing using simple top down parsing for the sentence “**The big dogs barked**” using the grammar given below:   S->NP VP  NP->ART N  NP->ART ADJ N  VP->V  VP->V NP | 10 | 3 | 2 | 1 | 1.7.1 |
| 12 | Discuss the role of Probabilistic Context-Free Grammars (PCFGs) in Natural Language Processing (NLP), specifically focusing on how they improve upon traditional context-free grammars in parsing ambiguous sentences. | 10 | 2 | 2 | 1 | 1.7.1 |
| 13 | Explain the concept of reference resolution in natural language processing and its importance in tasks such as coreference resolution and pronominal anaphora resolution. | 10 | 2 | 3 | 2 | 1.7.1 |
| 14 | Given a dataset of text documents, describe how you would use Word2Vec, Skip-gram, and GloVe algorithms to generate word embeddings. Include in your response the steps for preprocessing the data, choosing parameters for each model, and how each algorithm's generated embeddings could be used to enhance the performance of a text classification task. | 10 | 3 | 3 | 1 | 1.7.1 |
| 15 | Given a sample text, apply a Word Sense Disambiguation (WSD) method to identify the correct sense of ambiguous words. Describe the steps you took and the reasoning behind your choice of WSD method. | 10 | 2 | 3 | 1 | 1.7.1 |

**Course Outcome (CO) and Bloom’s level (BL) Coverage**

