|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |  |

****

**Presidency University**

**Bengaluru**

**Ph.D. Course Work End Term Examinations – JAN-FEB 2025**

**Date**: 06-02-2025

**Time**: 9.30 AMTO 12.30 PM

**Max Marks**: 100

**Weightage**: 50%

**Semester**:

**Course Code**: CSE863

**Course Name**: Advanced Natural Language Processing for Educational Applications

**School:** SOCSE

**Instructions:**

1. *Read the all questions carefully and answer accordingly.*
2. *Do not write any matter on the question paper other than roll number.*

**PART A**

**Answer all the Questions. Each question carries 10 marks. (6Qx 10M= 60M)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1.** | NLP has come a long way in the last decade. We initially had vectors like GloVe, FastText, and word2vec, before coming up with pre-trained language models in the late 2010s to finally large language models of today. Match each of the following models with their dimension sizes.   |  |  |  | | --- | --- | --- | | **Sl. No.** | **Model Name** | **Model Size** | | 1. | ALBERT (A Lite BERT) | Upto 300 dimensions | | 2. | BERT (Bidirectional Encoder Representation for Transformers) base | 12 million parameters | | 3. | BERT-large | 110 million parameters | | 4. | GloVe | 124 million parameters | | 5. | GPT2 (Generative Pre-trained Transformer) | 336 million parameters | |
| **2.** | Mention the term for each of the following definitions:  a. A point on the screen where the reader has focused their gaze for a reasonable period of time.  b. The areas of the screen which we need to collect gaze Behaviour from (like the area around words)  c. The movement of the eye from one point on the screen to another.  d. The movement of the eye from the current point on the screen to a later one.  e. The movement of the eye from the current point on the screen to an earlier one. |
| **3.** | Between 2016 and 2018, there were a number of systems that were described for holistic automatic essay grading. Some of them are Fei Dong and Yue Zhang’s 2016 system, Kaveh Taghipour and Hwee Tou Ng’s 2016 system, Fei Dong, Yue Zhang and Jie Yang’s 2017 system. Describe **any one** of the above systems. |
| **4.** | Mathias et al. (2020) which was published at IJCAI 2020, describes a number of tasks in NLP where gaze Behaviour is **learnt** (and not just recorded and used). Mention any 5 such NLP tasks, and any one paper which uses each of them. |
| **5.** | Explain the necessity of using Quadratic Weighted Kappa for automatic essay grading. |
| **6.** | Explain the differences between Automatic Essay Grading and Automatic Short Answer Scoring. |

**PART B**

**Answer all the Questions. Each question carries 20 marks. (2Qx 20M= 40M)**

|  |  |
| --- | --- |
| **7.** | The lexical simplification pipeline (described by Shardlow (2014)), consists of 5 steps, namely complex word identification, substitution generation, substitution selection, substitution ranking, and finally sentence simplification. Describe each of these steps for the example sentence “The cat perched on the mat”. |
| **8.** | At the TSAR 2022 shared task on lexical simplification substitution generation and ranking, a team from Presidency University participated. Their system consisted of 3 major steps, namely candidate token generation, candidate word selection, and candidate word pruning. Describe those steps for the sentence: “The weather predicted for Bengaluru is **torrential** rainfall for the next 5 days.” Here, the target word to be replaced is **torrential** (in bold face). |