NLP and Information Retrieval
Sample Exam 3

Generated by ChatGPT

Instructions

- Answer all questions.
- Show all calculations where required.
- Justify your answers clearly in transfer questions.

Task 1: Terminology & Basic Concepts

1a) (Knowledge)
Define the following terms in one or two sentences each:

- **Precision**
- **Recall**
- **F1 Score**

1b) (Understanding)
Why might **F1** be more informative than **Accuracy** in scenarios where the dataset is highly **imbalanced**?

Task 2: True/False – Foundations

Mark each statement **True (T)** or **False (F)** and provide **one sentence of explanation**:

1. Lemmatization always strips affixes down to the root form, regardless of context.
2. In a pipeline for text classification, feature extraction is optional if using a bag-of-words approach.
3. Boolean Retrieval evaluates partial matches by computing similarity scores for each document.
4. Kneser-Ney smoothing is a method used in language modeling to handle unseen n-grams.
5. A GloVe embedding can dynamically change depending on the words around it in a sentence.

Task 3: Naïve Bayes & Text Classification

3a) (Knowledge)
Briefly outline how a **Naïve Bayes classifier** estimates $P(\text{Class} \mid \text{Features})$. Why is the assumption called "naïve"?

3b) (Transfer)
You are performing **sentiment analysis** (positive vs. negative) on product reviews. Given a small training set, would you pick **Naïve Bayes** or a **Transformer-based** classifier? Justify your choice regarding **data size, complexity, and potential accuracy**.

Task 4: Inverted Index & IR

4a) (Knowledge)
Explain how an **inverted index** is constructed. Include the terms: posting list, document frequency (DF), and term frequency (TF).

4b) (Mini Exercise)
You have the following small corpus:

- **D1**: "Time flies like an arrow"
- **D2**: "Fruit flies like a banana"

Lemmatize and ignore function words ("a", "an"). Construct the **inverted index** for the terms "time", "fly", "like", "arrow", "fruit", "banana". Show the postings lists.

Task 5: BM25 Calculation (Short Exercise)

You have a collection of three documents (D1, D2, D3). Each term's frequency and document lengths are as follows:

- **D1** length = 4 tokens; term "cat" appears **2** times.
- **D2** length = 8 tokens; term "cat" appears **1** time.
- **D3** length = 8 tokens; term "cat" appears **0** times.

Assume average document length (**avgDL**) = 6. The query is "cat", and we use **BM25** with $k_1 = 1.2$ and $b = 0.75$.

Which document likely gets a **higher score** under BM25, D1 or D2? (No need for full formula numeric detail; a short reasoning step is enough.)

Task 6: Neural IR – Re-Ranking

6a) (Understanding)
In a **two-stage** IR pipeline with a **BERT re-ranker**, how are query and document tokens typically combined as input for BERT?

6b) (Transfer)
You have an **e-discovery** system for legal documents—**high recall** is mandatory. Explain why a **BM25 first stage** plus a **BERT re-ranker** might outperform a single-stage approach.

Task 7: Dense Retrieval & Vectors

7a) (Understanding)
What does **"dense retrieval"** mean, and how does it differ from **sparse** approaches (like BM25)?

7b) (Short Calculation)
Given a query embedding $\mathbf{q} = [0.6, 1.2]$ and two passage embeddings:

- $\mathbf{p}_1 = [1.0, 0.5]$
- $\mathbf{p}_2 = [0.0, 1.1]$

Compute the dot product **score** for each passage. Which passage ranks higher?

Task 8: Byte-Pair Encoding (BPE) & Subword Splits

8a) (Knowledge)
Why does **BPE** often outperform simple character-based tokenization in languages with rich morphology?

8b) (Do Stuff)
You have the initial subword vocabulary {_ (space), a, t, o, r}. The text is:

"a t a t o r a tor"

Show one or two hypothetical merge steps and how they reduce the total token count.

Task 9: Instruction Tuning vs. Prompting

9a) (Understanding)
Explain how **instruction tuning** changes a model's behavior differently than simply providing a long prompt.

9b) (Transfer)
Your **FAQ chatbot** must handle a variety of user instructions. Outline **two** ways instruction tuning can help produce better responses and mention **one limitation**.

Task 10: RLHF & Model Alignment

10a) (Understanding)
In **RLHF**, what is the role of the **reward model**, and why do we prefer **pairwise** human feedback?

Task 11: Handling Long Context in Transformers

11a) (Knowledge)
What problem arises in **self-attention** when handling long sequences?

11b) (Transfer)
You need to process **25,000-token transcripts**. Would you use **sparse-attention Transformers** or **retrieval-augmented generation (RAG)**? Justify.