# Exercise Sheet 9

## Conditional Random Fields and IR

**Deadline: 18.06.2025 23:59**

**Guidelines:** You are expected to work in a group of 2-3 students. While submitting the assignments, please make sure to include the following information for all our teammates in each of your PDF files/python scripts:

**Name:**

**Student ID (matriculation number):**

**Email:**

Your submissions should be zipped as **Name1_id1_Name2_id2_Name3_id3.zip** when you have multiple files. For assignments where you are submitting a single file, use the **same naming convention** without creating a zip. For any clarification, please reach out to us on the **CMS Forum**. These instructions are mandatory. If you are not following them, tutors can decide not to correct your exercise.

**Exercise 9.1 - HMMs and CRFs**                                   (1+1=2 points)

   a) Consider the sentence *"Will you book the flight?"*. Assume you are designing a sequence model for POS tagging.

      i) Explain whether a first-order Hidden Markov Model (HMM) is sufficient for resolving such ambiguities.
         Support your answer using properties of HMMs.                    (0.5 points)

      ii) Propose an alternative graphical model that can capture such long-range dependencies more effectively.
         Explain how it addresses the limitations of HMMs.                (0.5 points)

   b) In the context of Conditional Random Fields (CRFs), cliques play an important role in defining the factorization of the probability distribution.

      i) Define what is a clique and what are maximal cliques.             (0.5 points)

      ii) Given the graph in Fig. 1, list all cliques and identify the maximal clique(s). Then, explain how maximal cliques determine the factorization of the joint distribution in a CRF.                                                          (0.5 points)
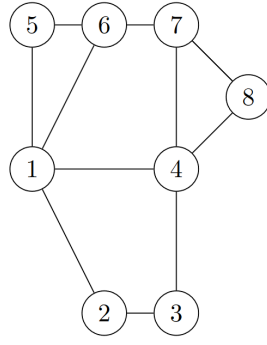
Figure 1: Undirected graph

## Exercise 9.2 - Information Retreival                    (1.5+1.5=3 points)

a) Given the following results for a query $Q$:

| Document | Rank ($k$) | Relevance |
|----------|------------|-----------|
| A | 8 | 1 |
| B | 1 | 1 |
| C | 2 | 1 |
| D | 3 | 0 |
| E | 9 | 0 |
| F | 4 | 1 |
| G | 7 | 1 |
| H | 6 | 0 |
| I | 5 | 1 |
| J | 10 | 1 |

Table 1: Search results for query $Q$

Each document is associated with:

- A **Rank** ($k$), which indicates the position in the retrieved list (lower is better),
- A **Relevance** score, where 1 indicates relevant and 0 indicates non-relevant.

**Tasks:**

   i) Compute the **Recall@$k$**.                          (0.5 points)
   ii) Compute the **Precision@$k$**.                      (0.5 points)
   iii) Compute the **Average Precision (AP)**.            (0.25 points)
   iv) Compute the **Mean Reciprocal Rank (MRR)**.        (0.25 points)

b) Fill in Table 2 with the correct metric to be used in each case.     (0.25 points each)

## Exercise 9.3 - Transformers and CRFs                    (3.5+1.5=5 points)

See attached notebook

2

| Task | Metric |
|---|---|
| Filtering emails to prevent spam while avoiding blocking important legitimate messages, where minimizing false alarms is crucial to user experience. | |
| Screening patients for serious diseases where failing to identify any true cases could have severe consequences, even if it means more follow-up tests. | |
| Detecting fraudulent transactions in financial systems where both false accusations and missed fraud cases have costly implications, especially with highly imbalanced data. | |
| Ranking search engine results by relevance, ensuring that highly relevant documents appear near the top to satisfy users' diverse information needs. | |
| Evaluating an autonomous vehicle's ability to accurately detect multiple types of objects (pedestrians, vehicles, signs) consistently across all categories. | |
| Providing quick, accurate responses in a conversational assistant, where the position of the first relevant answer greatly affects user satisfaction. | |

Table 2: Metrics table