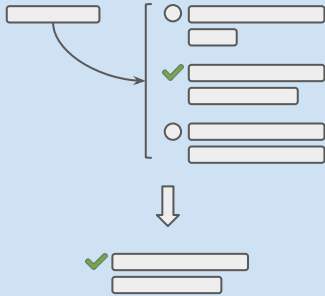# 4 Projects

1. Text Retrieval
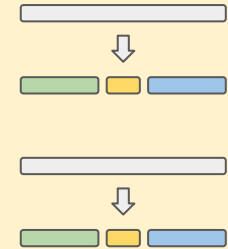
2. Recommender systems
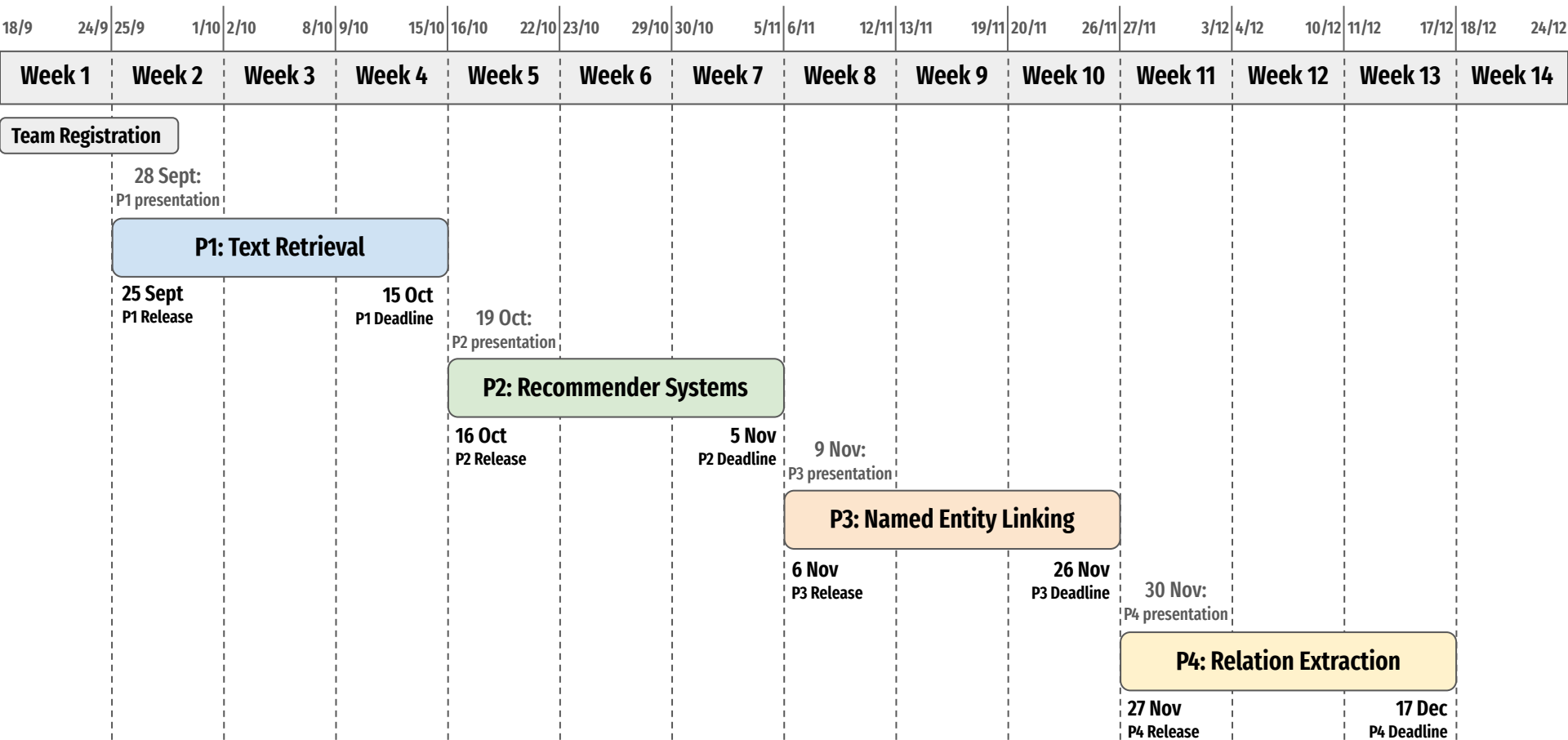
3. Named Entity Linking
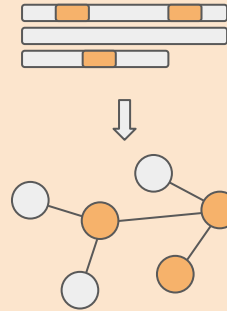
4. Relation Extraction

Timeline

18/9	24/9 25/9	1/10 2/10	8/10 9/10	15/10 16/10	22/10 23/10	29/10 30/10	5/11 6/11	12/11 13/11	19/11 20/11	26/11 27/11	3/12 4/12	10/12 11/12	17/12 18/12	24/12
Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14

Team Registration

28 Sept:
P1 presentation

P1: Text Retrieval

25 Sept
P1 Release

15 Oct
P1 Deadline

19 Oct:
P2 presentation

P2: Recommender Systems

16 Oct
P2 Release

5 Nov
P2 Deadline

9 Nov:
P3 presentation

P3: Named Entity Linking

6 Nov
P3 Release

26 Nov
P3 Deadline

30 Nov:
P4 presentation

P4: Relation Extraction

27 Nov
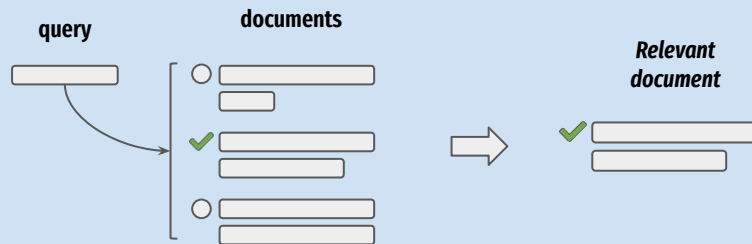P4 Release

17 Dec
P4 Deadline

1. Text retrieval

DESCRIPTION

In this project we need to rank documents based on their relevance to a query, where documents can be retrieved from the full document collection provided.

It models a scenario where you are building an end-to-end retrieval system.

EVALUATION

- hits@1, hits@5, hits@10

2. Movie Recommendations

DESCRIPTION

Recommender Systems (RS) are algorithms that aim to suggest relevant items to users. RS are widely used in many applications such as e-commerce, social media, and entertainment.

The goal of this project is to **develop a recommender system that is able to predict the movie rating of a user for a given movie**.

EVALUATION

- RMSE / MSE / MAE

user past movie ratings

New movie unrated by the user

Estimated user rating of the new movie

3. Named Entity Linking

DESCRIPTION

Named Entity Disambiguation (NED) is the task of mapping entities (e.g. person, location, companies), from an input text document to corresponding unique entities in a target Knowledge Base (KB).

You have to develop a **NED model trained on AIDA-CoNLL dataset** that is able to link the entities in a given input document to a matched unique entity in YAGO.

EVALUATION

- Accuracy / Precision / Recall / F1
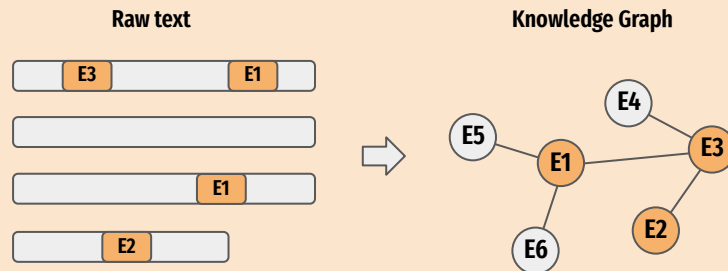
4. Relation Extraction

DESCRIPTION

In this project, we propose to tackle the Relation Extraction (RE) task. RE involves extracting the relation between two given entities based on their related context.

You have to develop a **RE model trained on the annotated dataset** where each sample contains two named entities (persons or organisations) and a relation type between them. Using only the context, the goal is to label the relation between the two given entities among *41 relation types* or to label it as *no relation* if there is no defined relation between the two.

EVALUATION

- Accuracy / Precision / Recall / F1

Raw text → Relation extraction (head, rel, tail)
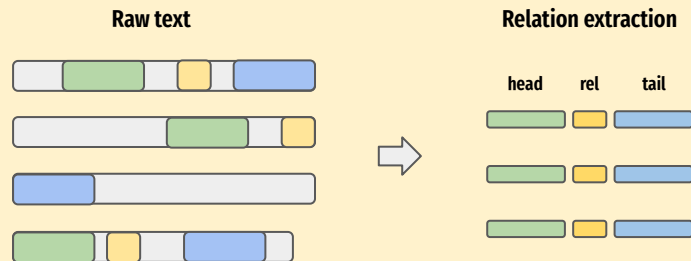
Logistics

PROJECT EVALUATION

Projects count 50% of the final grade.

- <u>40%</u> : Results - Metrics   [Comparison with baselines]

- <u>30%</u> : Code
  - Working code (20%)
  - Code quality and documentation (10%)

- <u>30%</u> : 2-page Report   [Moodle submission]
  - Originality of approach (10%)
  - Interpretation of results (10%)
  - Report presentation & clarity (10%)

LATE DAYS POLICY

- 5 days in total for all Projects.

INFRASTRUCTURE / IMPLEMENTATION

- Python
- Notebooks / .py scripts
- Locally / Colab / Kaggle

KAGGLE

- Kaggle dashboard for each Project
- Run code + submit predictions

Team Registration

Teams of 2 or 3 people

Registration deadline 28 Sept

https://forms.gle/VFN7Xif39f89nttX9