Project Specification

The project involves conducting an IR experiment using advanced methods (e.g., Transformers). The experiment must be in the scope of information retrieval, containing evaluations and some advanced concepts covered in class. The project consists of four parts, graded separately:

- **Design Document** - Sketching your project idea.
- **Project** - Your implementation of an advanced IR project, done in groups.
- **Presentation** - presentation of your work in class, done with your group.
- **Report** - a description of your system, your experiments and findings in the form of a scientific paper.

The project will be conducted in groups of 4 people.

Programming Language: Python

Important Dates

- Group Formation until: 15.10.2024 23:59 - has been extended!
- Project Handout: 29.10.2024 (in lecture)
- Design Document due: 03.12.2024 23:59
- Project Submission due: 12.01.2025 23:59
- Project Presentation due: 12.01.2025 23:59
- Project Presentations: January lecture units (14.01., 21.01., 28.01.) - slots need to be booked in TC

Design Document (15p)

1 document per group

The document should sketch the general idea in four A4 pages (Drawing/Diagram + Writing) and contain:

- Title
- Group Number
- Members + tentative roles (= responsibilities)

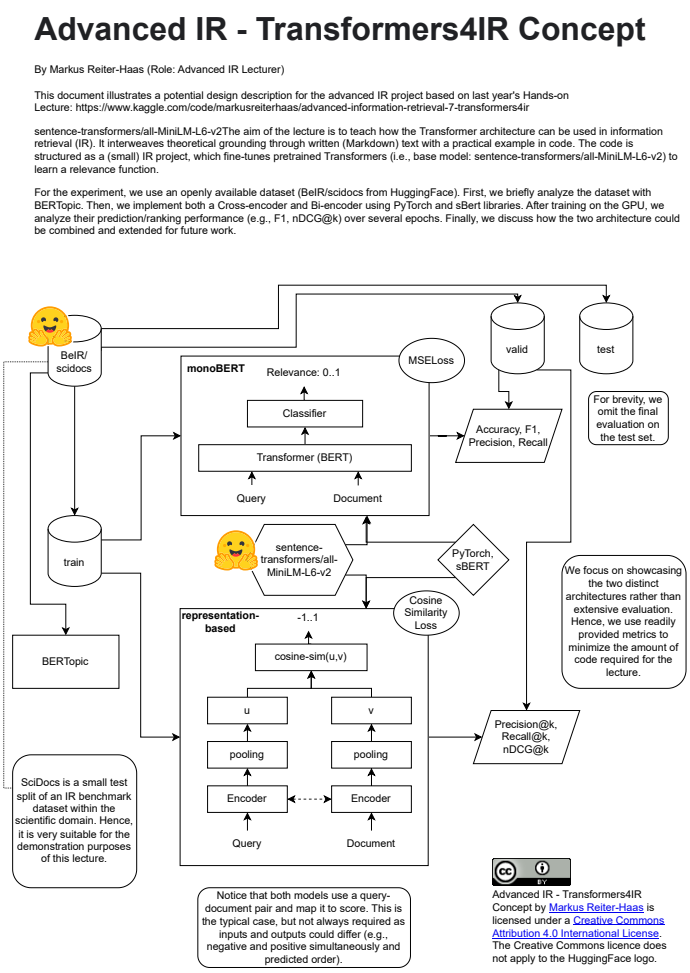

Figure 1: Example Sketch from WS22

- Abstract
- Visual Depiction - see an example in Fig. 1

The design document must clarify:

- **Idea** (goal, e.g., problem to solve formulated as research question)
- **Main Task** (focus of your work)
- **Dataset + Processing** (resources used)

- **Methods/Models** (how you intend to solve the problem)
- **Evaluation** (how results are measured)

Submission as PDF in TeachCenter.

Project (30p)

The project should be developed with your group and must be an advanced IR system. Submit the code in advance and provide a link to your code repository in TC.

Repository

The repository should be public (e.g., GitHub, Zenodo): **Be proud of your work.** If not possible, contact us at least 2 weeks in advance to share the code privately.

Must include:

- Code for reproducibility
- Plots (either separate or within Jupyter notebooks)
- Brief dataset description (e.g., location)
- Optionally: the dataset itself

Top projects will be highlighted on the lab website: https://socialcomplab.github.io/advancedIR-2023-showcase/

Presentation (15p)

Final duration of presentation will be announced in TC. Submit slides as PDF in TC.

Presentation Content

- Title slide
- Introduction (RQ/motivation)
- Data + Methods (+ optional theoretical background)
- Results (+ analysis/interpretation)
- Conclusion (incl. limitations/biases)

Title page must include:

- Name of the project (defined by group)
- Group number
- Group members + team roles
- Link to repository

Report (20p)

Write a report on your experiments as a scientific paper (max. 4 pages, 12pt font, excluding title page). Submit as PDF in TC. The structure should be:

- **Title page** with group number and names
- **Introduction**: Motivation and research question
- **Related Work**: Brief literature or project discussion
- **Experiments and Results**: Conducted experiments and findings
- **Conclusion**: Key takeaways and future work

Tips

An important aspect of an IR experiment is the analysis and evaluation of IR systems/models. For instance, an image generator is not IR-related and its evaluation may be unsuitable for this course.

Organization:

- Start project work before handing in the design document. Test if design components (e.g., dataset) are feasible.
- Use feedback from the design document to improve the project.
- Seek help early if needed (e.g., in TeachCenter).

Programming:

- Jupyter is recommended, but for long tasks, scripts may be better.
- Use Numpy or PyTorch instead of plain Python.
- For large datasets, create a smaller subset and report statistics in the presentation.
- Attend lectures, especially on Transformers.

Presentation:

- Stick to time. Rehearse beforehand.
- Engage with the audience and observe time signals.
- Nervousness is normal; preparation is key.

Grading Considerations:

- A clean setup is crucial, though code style is not graded.
- The setup must avoid issues like data leakage.
- Results require analysis/interpretation beyond mere presentation.
- Negative results are valid but must be carefully analyzed and validated. For example, a high-performing random model indicates a methodological issue, not a valid negative outcome.