486 Information Retrieval and Web Search

Term Project
Winter 2025

The project is an open-ended exploration into some aspect of information retrieval and web search. You will have to work in teams of five, and design and implement a system that performs a task relevant to the broad field of Information Retrieval and Web Search, which addresses a socially relevant problem.

The project will be graded based on the quality of your technical solution to the problem, the execution (data collection, system implementation and evaluation), and the presentation (written report and oral and poster presentation).

Project requirements

The project accounts for 30% of the final grade. In the points scheme used for the class (100 points for each 10%), this means the project is worth 300 points. It also means that a team of five will account for 150% of the final grade, so the amount of effort expected for the project is roughly equivalent to the amount of effort one would put into 1.5 semester-long class.

There are three checkpoints, and five deliverables associated with the project:
02/18, 03/12

- Project checkpoint 1 (30 points): 02/07: Project proposals submitted via the Google form provided on Canvas. Only one form per team, please.

- Project proposal presentation (30 points): 02/21: Project proposals presented in person in Tishman Hall, 9-11am.

- Project checkpoint 2 (50 points): 02/27: Intermediate project report, submitted via Canvas.

- Project checkpoint 3 for peer review (25 points): 04/04: Complete draft of the final project report, submitted via Canvas.

- Peer project reviews (15 points): 04/10: Comprehensive review of one peer project

- Final project report (60 points): 04/17: Final project report, submitted via Canvas.

- System and datasets (60 points): 04/17: Code and data, submitted via Canvas.

- Project presentation (30 points): 04/18: Posters presented in person during the final EECS 486 Project Presentations session, Tishman Hall, 9-12pm. Posters should also be submitted on Canvas by 04/18.

The same policy of a maximum three days delay with a 10% penalty for each day late apply for most of these project deliverables; the only exception are the project proposals and final project presentations, which can only happen in person.

Only one person per team needs to submit on Canvas.

For all the intermediate reports and for the final report, please use the ACL templates (either Word or Latex) https://github.com/acl-org/acl-style-files

1   Project checkpoint 1

Submit your project via the Google form available on Canvas.

- The social problem your project will solve.

- The task definition (the potential technical solution to the problem).

- Initial ideas on datasets to use (existing datasets, or datasets that need to be compiled).

- Potential evaluation metrics.

- The members of the team; exactly 5 members, no exceptions.

2   Project proposal presentations

You will have to make an in-person presentation of your project. You will present your project idea in person, during a 2 hour session to be held in person on 02/21, 9-11am, in Tishman Hall. Poster boards of size 22 x 28 in will be available. You are not expected to print a full-size poster; instead, highlight the main ideas on 2-3 US Letter sheets of paper. Among others, this should include:

- Title of your project

- The social problem you are solving and the potential impact.

- The technical solution you are proposing to address the problem.

- Any already known steps: algorithms to be used, candidate datasets, evaluation, etc.

In addition, please also add one 1/3-1/2 US Letter for each team member, including: name, a picture, one thing Google does not know about you (and you are willing to share)

Project proposal presentations will receive feedback from the instructors and from your peers.

3   Project checkpoint 2

A project report that includes:

- Project description (at least 1 page), including the social problem you are addressing and its impact, and the technical solution you are proposing.

- Related work, with at least three references (at least 1/2 page)

- Data collection method description, data annotation method, interesting data samples (as appropriate) (at least 1/2 page).

- Method description, including evaluation methodology (at least 1 page)

4   Project checkpoint 3

A close-to-final project report that includes:

- Project description, including the social problem you are addressing and its impact, and the technical solution you are proposing. Expand the previous version with stronger arguments, examples, etc.

- A review of related work, and how it relates to your project. Aim to include at least six references (more are encouraged).

- A detailed description of the datasets you use, including a description of the method you used to collect the data, issues encountered during data collection, data annotation (if any), interesting/representative examples from the data, etc.

- A detailed description of your approach, highlighting any original contributions.

- Experiments and results, with a description of the evaluation methodology (metrics, baselines), the experiments you ran, presentation of results, comparison with baselines and alternative methods, discussion of results (accuracy, but also issues related to efficiency, scalability, etc.)

- Conclusions, main contributions of your project, how your solution can solve the social problem you are addressing, what worked and what did not work, considerations for future work.

5   Peer project reviews

A detailed review of one project from your peers, including:

- A summary of the problem addressed and the solution proposed.

- A list of project strengths (aim for at least three strengths, each addressed in detail)

- A list of project weaknesses, along with suggestions for how to address these weaknesses (aim for at least three weaknesses, each addressed in detail).

- A score (on a scale of 1-10), with 10 being "excellent project" and 1 being "very weak project"

6   Final report

The final report should include a description of the complete work on the project, explicitly accounting for the reviews you received. It should include all the sections from Project checkpoint 3 with significant improvements based on the reviews, as well as the following two sections:

- A section on how you addressed the reviewers comments (at least 1/2 page)

- Description of the individual contributions of each team member; please include one paragraph per team member.

The final report should be at least six pages and at most ten pages long.

7   Software and datasets

Your project will have to include a complete implementation of your approach. The software will have to be written in Python, and it will have to run on a Linux platform. You can use external libraries as needed (provided they do not make the project trivial). The grade for this part will be based on the quality of your implementation (which includes code documentation and a complete README file). Please also include all the datasets used in the project, including raw and annotated data (depending on the project).

8   Poster presentation

You will have to make a poster presentation of your project. You will have to present your poster during a 3 hour poster session, to be held on 04/18, 9-12pm, in Tishman Hall. Demos are also encouraged (but not mandatory); if your project includes a demo, it should be available for presentation during the poster session.

9   Important notes

- All teams have to have exactly five members.

- All team members have to **equally** contribute to **all** the project components.

- The project has to be written in Python.

- The use of libraries is allowed to the extent they do not make the project trivial. If in doubt, ask the instructors.

- The grading of the project is not based on accuracy. The exploration of new, interesting ideas (even if they turn out not to work) is appreciated more than incremental improvements on old ideas.