The project is an open-ended exploration into some aspect of information retrieval and web search. You will have to work in teams of four, and design and implement a system that performs a task relevant to the broad field of IR. The project will be graded based on the quality of your solution to the problem, the execution (data collection, system implementation and evaluation), and the presentation (written report and oral presentation).

1 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. There are four checkpoints, and three deliverables associated with the project

- Project checkpoint 1 (25 points): Due 02/27, submitted via CTools.
- Project checkpoint 2 (25 points): Due 03/25, submitted via CTools.
- Project checkpoint 3 (25 points): Due 04/03, submitted via CTools.
- Project checkpoint 4 (0 points): Due 04/17, submitted via CTools.  
  *Note on 04/15: Project checkpoint 4 has been removed. The points have been redistributed.*
- Final report (85 points): Due 04/21, submitted via CTools.
- System and datasets (110 points): Due 04/21, submitted via CTools.
- Project presentation (30 points): Due April 15/20, in class.

The same policy of a maximum three days delay with a 10% penalty for each day late apply for all these project deliverables.

Only one person per team needs to submit on CTools.

For all the intermediate reports and for the final report, please use the NAACL template (either Word or Latex). If necessary (e.g., because of editing the report under Google Documents) it is acceptable to use a single-column format. [http://naacl.org/naacl-pubs/](http://naacl.org/naacl-pubs/)

1.1 Project checkpoint 1

A project report that includes:

- Project description (at least 1 page)
- Data collection method description, data annotation method, interesting data samples (as appropriate) (at least 1 page). Statistics on the data collected to date.

Also: (partial) data collection/annotation.
1.2 Project checkpoint 2
An updated project report that also includes:
- Related work, with at least 4 references (at least 1 page)
- Updated data collection method description, data annotation method. Statistics on the final dataset.
Also: completed data collection, including annotations as needed.

1.3 Project checkpoint 3
An updated project report that also includes:
- Method description, including evaluation methodology (at least 2 pages)
Also: partial method implementation.

1.4 Project checkpoint 4
An updated project report that also includes:
- Updated method description.
- Evaluation of the method with results and analysis of results (at least 1 page)
Also: complete method implementation.

1.5 Final report
The final report should include a description of the complete work on the project. It should include:
- A description of the problem you are addressing, including definition, examples, motivation, applications.
- 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, what worked and what did not work, considerations for future work.
- Description of the individual contributions of each team member.

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

1.6 System 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).
1.7 Oral presentation

You will have to present your project in class. The presentation will last for 8 minutes, with an additional 2 minutes for questions. Demos are also encouraged (but not mandatory); if your project includes a demo, please allow for some time during your presentation to show the demo.