CS458 - Final Project
Fall 2012

1 Overview

Throughout the first two-thirds of this course we have learned about the foundational concepts required to build a modern information retrieval system. Through our assignments, we have constructed the framework for a basic IR system. In this project, you'll be adding the bells and whistles. It is a chance for you to learn more about a sub-field of IR and to explore the experimentation process in CS. Not only will you be implementing an addition to our system, you will be evaluating your addition and doing a final write-up and presentation.

This project may differ from many of your other class projects that you've had before because your project will be not be viewed in isolation. My hope is that with good planning, we will be able to integrate all of these projects together into one (yet to be named) IR system. Besides having something interesting you can show other people about your hard work, this will also give you some exposure into the challenges of working on a modestly sized software project (for those that didn't take software development with me :).

The major implication of this will be that planning for this project will be **very** important since your project will interact with other people's projects. We will all be building from the same starting code base, and so you will need to specify any interactions that you will have with our current code.

2 Logistics

Schedule

date	description
11/11	Form team and e-mail me project topic
11/13	Project proposal draft and in-class discussion/planning
11/15	Finalized project proposal
11/27	Status report
12/7	Paper draft
	initial code submission
12/9	Reviews
12/14	Final paper
	Presentation
	final code submission

Project requirements

The project:

– must be related to something we have discussed (or will discuss) in class

– must integrate with our current system, though, it is likely we'll need to make changes to interface properly

– must involve some sort of evaluation

– should work out to around 20 hours of coding work per person (this isn't a lot, so plan accordingly)

Project ideas

– GUI and/or backend

– Snippet/summary generation

– Index compression

– Distributed indexing

– Faster/approximate ranked retrieval

– Spelling correction/Query suggestion

– Relevance feedback

– Audio retrieval

– Image retrieval

– Clustered search results

– Crawling (though come talk to me beforehand if you're thinking about doing this)

3 Non-coding work

Much of your time for this project will be spent on coding up your addition to the system, however, there are a handful of other things you will be required to do along the way.

Project proposal

Your first task will be to come up with your project group and figure out what you'd like to work on. You'll need to e-mail this to me first. I'll try and respond quickly if I see any issues with your topic.

Once you've decided got the topic and team figured out, you'll need to write a draft of the project proposal. The day the proposal drafts are due, we will spend the entire class talking about the different projects and how they will all interact, so it's important that you spend some time thinking about this before hand.

Based on our discussion in class, you will then go back and revise any issues and submit the final version.

Your project proposal must include the following information:

- Members of the team. I'm *strongly* encouraging groups of 2 or 3. If you want to work solo, please come talk to me.

- A 1-2 paragraph description of your project including:

  – What you hope to accomplish.

  – How you will evaluate your approach. You should think about what would be an appropriate way for evaluating your success. For example, for index compression you could measure the size reduction, for distributed indexing you could measure the speed increase and for spelling correction you could measure retrieval performance improvement for mispelled queries.

  – What resources you will use/need including code, data, etc. All of these resources MUST be obtained by the time the final draft of the proposal is due. Come talk to me (early) if you're having trouble finding appropriate data.

- One or two research papers from the literature that tackle the same problem. Google scholar is a good place to start, though I can also help you if you're having trouble finding papers.

- Logistics:

– How you will integrate with the current code base including any changes that need to be made to existing code.

– Things to be accomplished by the status report. These don't need to be very detailed, but should be very concrete.

Status reports

In your project proposal you will specify a list of intermediary goals/accomplishments. A status report is a one paragraph description of the current status of the project, including:

– to what extent you accomplished your proposed tasks

– any problems or issues that have arisen

– the number of hours each person put in the project

This will also be an opportunity to communicate to me if there are any changes to your specification, etc.

Final Paper

Your group will also need to put together a final paper to discuss what you did and your results. You should view this paper as a mini-research paper. We will use the ACM paper format which is used by SIGIR:

> http://www.acm.org/sigs/publications/proceedings-templates

The website has templates for both latex and word, either of which are fine. You paper should be *no more than 2 pages long.* Even though it is short, I expect it to be well written, well organized and present what you've done (including your results) clearly and concisely.

You may organize the paper however you like, but a common approach would include the following:

- **Abstract:** A very high-level view of the problem, approach and results. An abstract is almost never more than a paragraph.

- **Introduction:** Describe the problem and motivate why the problem is interesting/useful.

- **Algorithm description:** Clearly describe your algorithm including any challenges you encountered.

- **Results:** Describe your data, experimental setup, evaluation criterion and how well your system performed. You should spend some time discussing the results, including anything results that are surprising or interesting.

- **Conclusion:** A brief summary of the paper including any challenges, where to next and any high-level comments you have at the end of the project.

Reviews

Each of you will be asked to review two papers written by other groups in your class. I'll talk more about good reviews as we get closer. For these reviews it will be important that you give very specific feedback about things that were good/bad/unclear in the paper since the authors will use your feedback to improve their final paper.

Presentation

Each group will give a short (15-20 minutes) presentation of your work during our final exam period. At a high-level, your presentation will have a similar flow to your paper. Your presentation must include the following information:

- **Problem**

- **Motivation:** Why is what you did useful?

- **Approach:** How did you solve the problem?

- **Results:** Evaluation? How well does it work?

4   Coding and Code submission
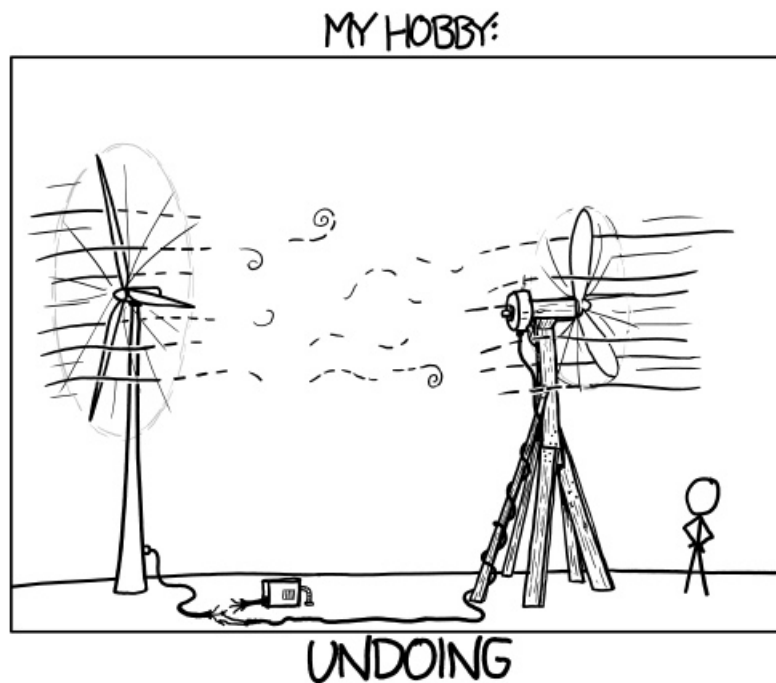
Since we are developing a system as a class, you will need to submit your final working code. I will find a home for our project and we'll talk about integration, etc. We will likely use Git to create a shared repository so you guys can coordinate to make a final working system.

When you submit your paper draft, you will also submit/check-in a working/compiling version of your code so that we can work out any last integration issues. When you submit the final draft, you will submit/check-in a final version of your code that should address any last integration problems that have been identified.

5 Grading

The project will represent a majority of your work for the last third of the class, so don't get too concerned that there are a lot of things to do. You'll have about 3.5 weeks to accomplish the tasks, which should be plenty of time if you stay on top of things and put in regular work.

- Group and topic e-mail (*2 points*)
- Project proposal draft (*5 points*) - Did you think about the project beforehand and have something reasonable written?
- Project proposal (*20 points*) - Meets specifications above.
- Status reports (*15 points*) - Meets specification above. In addition, I will also be grading your group based on the amount of progress you've made and the amount of effort put in. I don't want you leaving all of the work until the end of the semester!
- Paper draft (*10 points*) - Do you have a reasonable draft in place for other students to review?
- Initial code submission (*10 points*) - Compiles, follows integration specifications
- Paper reviews (*20 points*)
- Final paper (*100 points*) - 50 points: meets specifications above. 50 points, overall progress made on the project and quality of work.
- Presentation (*20 points*)
    - Covered content
    - Organized and well-prepared
    - Presentation style
- Final code submission (*10 points*) - Compiles and all integration issues have been worked out.