1 Meeting Times

• Lectures Monday, Tuesday, and Thursday, 3:00pm to 4:15pm in 1304 Siebel; the first day of class is June 2
• Midterm exam in class July 24
• Project presentations in extended class (3:00pm - 5:00pm) on August 4 and August 5
• No class after project presentations!

2 Communication

2.1 Course Web Site

All course information is available on the web site: http://web.engr.illinois.edu/~massung1/su14-cs410/. When labs and MPs are released, their instructions will be detailed here. Course announcements will be made on the home page with relevant links included.

2.2 Piazza

The most important forum for communicating in this class is the course's Piazza. Piazza is like a newsgroup or forum—you are encouraged to use it to ask questions, request clarifications, express opinions, and give advice.

The Piazza site for this class is https://piazza.com/class#summer2014/cs410. You are welcome to sign up, and you can do so directly if you use your netid@illinois.edu email address.

We expect that you will be courteous and post only material that is somehow related to CS 410 (however slightly). The posts will be lightly moderated.

Note that private posts to Piazza can be used for things like conflict requests, or for letting us know that you have that sinking feeling—anything you don’t really want to share with your classmates.

3 Prerequisites

We assume you have taken CS 225 and have a good working familiarity with linux. You will need to be familiar with common linux commands such as ssh, scp, and cat. Using the vim editor will be useful when remotely connecting. We will submit assignments with svn.

Significant programming experience from CS 225 will be helpful so you can focus more on the methods being explored rather than the syntax. You will probably use a scripting language for MP 1,
Java for MPs 2, 3, and 4, and potentially C++ for MP 5. Of course, you will also have to do extensive programming for your final project.

4 Textbooks

There is no official textbook for this course and no assigned readings. However, we do recommend the following resources.


More resources are on the course resources page: http://web.engr.illinois.edu/~massung1/su14-cs410/resources.html.

5 Grading

5.1 Graded Work

All grades will be published online. You can view your grades by visiting https://chara.cs.illinois.edu/gb/ and logging in.

Please check that your grades are recorded correctly. If you have any concerns, feel free to post a private question on Piazza. Do not wait until the last week of the semester to inform us of a grading mistake!

5.2 Point Breakdown

We will determine final course grades as follows:

- **3 credit hours**
  - 33.3% MPs
  - 33.3% Midterm exam
  - 33.3% Final project

- **4 credit hours**
  - 25% MPs
  - 25% Midterm exam
  - 25% Literature Survey
  - 25% Final project

After each assignment is graded, we will update your current grades. Histograms of grade distributions will also be released after each assignment and exam in the Chara gradebook.
5.3 Grade Cutoffs

We will use the standard grade cutoff points:

<table>
<thead>
<tr>
<th>Letter Grade</th>
<th>Point Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>[93, 100]</td>
</tr>
<tr>
<td>A-</td>
<td>[90, 93]</td>
</tr>
<tr>
<td>B+</td>
<td>[87, 90]</td>
</tr>
<tr>
<td>B</td>
<td>[83, 87]</td>
</tr>
<tr>
<td>B-</td>
<td>[80, 83]</td>
</tr>
<tr>
<td>C+</td>
<td>[77, 80]</td>
</tr>
<tr>
<td>C</td>
<td>[73, 77]</td>
</tr>
<tr>
<td>C-</td>
<td>[70, 73]</td>
</tr>
<tr>
<td>D+</td>
<td>[67, 70]</td>
</tr>
<tr>
<td>D</td>
<td>[63, 67]</td>
</tr>
<tr>
<td>D-</td>
<td>[60, 63]</td>
</tr>
<tr>
<td>F</td>
<td>(−∞, 60)</td>
</tr>
</tbody>
</table>

The letter grade cutoffs may be lowered, but they will not be raised. Do not count on the scale being adjusted at the end of the semester.

6 MPs

In order to get familiar with text mining and information retrieval software, there are several programming assignments. The first three are individual assignments and the last two are group-based.

<table>
<thead>
<tr>
<th>MP</th>
<th>Released</th>
<th>Due</th>
<th>Group?</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>June 6</td>
<td>June 15</td>
<td>No</td>
<td>Getting familiar with text</td>
</tr>
<tr>
<td>2</td>
<td>June 15</td>
<td>June 22</td>
<td>No</td>
<td>Using Lucene(^1) retrieval functions</td>
</tr>
<tr>
<td>3</td>
<td>June 22</td>
<td>June 29</td>
<td>No</td>
<td>Crawling; retrieval function competition</td>
</tr>
<tr>
<td>4</td>
<td>June 29</td>
<td>July 6</td>
<td>Yes</td>
<td>MapReduce</td>
</tr>
<tr>
<td>5</td>
<td>July 6</td>
<td>July 13</td>
<td>Yes</td>
<td>Using MeTA(^2)</td>
</tr>
</tbody>
</table>

After the last MP, you should focus most of your time working on the final project and preparing for the midterm exam.

7 Midterm Exam

There will be one midterm exam in this course; there is no final exam. The midterm will be on July 24 in class. Online students will be provided with their exams. The midterm exam will cover content from the first seven weeks of class; this will mainly concern (but is not strictly limited to):

- Text information systems overview
- Vector space retrieval models
- Language model retrieval models

\(^1\)http://lucene.apache.org/
\(^2\)http://meta-toolkit.github.io/meta/
- IR system implementation
- Evaluation in IR
- Feedback in IR
- Web search
- Applications: clustering, categorization, filtering, and recommendation

A detailed list of topics to study will be released on the course site before the exam. No cheat sheets or any other aids will be allowed during the exam.

8 Final Project

The course project is to give the students hands-on experience on developing some novel information retrieval and/or text mining tools. The project thus emphasizes applied research and “deliverables”, meaning that the outcome of your project should be something tangible, typically some kind of prototype system that can be demonstrated, though theoretical research projects are also fine. Group work is strongly encouraged, but not required.

There are many details about the project on the course site, but it consists of these main parts:

- 10% **Project proposal**, due June 13 by midnight
- 20% **Midterm report**, due July 3 by midnight
- 40% **Project report**, due August 2 by midnight
- 30% **Project presentation**, August 4 and August 5 in class

The proposal and midterm report will be graded pass/fail, and the presentation and final report will be graded with a rubric (to be released).

9 Literature Survey

Students taking CS 410 for 4 credit hours will be responsible for writing a literature survey on a topic of their choosing (related to this class’s material). The topic will be selected by the student with approval of the instructor. Often the selected topic is related to the course project that the student is involved in, but it does not have to be.

There are two main milestones for the literature survey:

- 10% **Survey proposal**, due June 26 by midnight
- 90% **Survey report**, due July 11 by midnight

The proposal will be graded pass/fail, and the presentation and final report will be graded with a rubric (to be released).
10 Resources

10.1 Computing

Since you are taking a class in the College of Engineering, you should have an account on the EWS (Engineering Workstation) machines. These machines are located in labs around campus, and have the necessary software installed for taking this course (which isn't that much). Their Web site is http://it.engineering.illinois.edu/ews/. It shows locations and current lab utilization.

If you want to use the EWS machines remotely, check out PuTTY and NXClient. Of course, you can also just use ssh to connect to a terminal.

There is also a virtual machine here https://wiki.engr.illinois.edu/display/cs225/Virtual+Machine that mimics the configuration of the EWS machines. However, in the Summer session we can't devote too much time to helping you get it set up (or debugging any issues).

10.2 Class calendar

All the dates and course meeting times mentioned in the syllabus are entered in the class calendar: http://web.engr.illinois.edu/~massung1/su14-cs410/calendar.html. Office hours will also be added.

11 Academic Integrity

You know the drill: do not cheat. On solo MPs (1, 2, and 3), work alone (but use Piazza and course staff!). In MP 4 and 5, you may work in a group of up to three students total. All code and text submitted needs to have been written this summer, after the MP has been released, by either you or your group (if applicable). Anything else is cheating. When working in a group, all group members need to submit files.

First-offense cheaters receive a zero on the assignment. It is mandatory for the instructor to report cheating incidents to the College of Engineering; this is not negotiable. Second-offense cheaters fail the course.

12 Acknowledgements

Thanks to Professor ChengXiang Zhai; some text borrowed from his course site for CS 410.

Thanks to you for reading the entire syllabus. Hopefully it makes your experience a bit easier and less stressful. An online copy of this syllabus can be found on the course site.