SI760/EECS597/LING702 Language and Information

Instructor: Dragomir R. Radev
University of Michigan
Fall 2002

Final project

Assigned: October 11, 2002
Proposal due: October 24, 2002
Project due: December 5, 2002
Project presentations: December 6, 2002

1 Introduction

Each student is responsible for designing and completing a research project that demonstrates the ability to use concepts from the class in addressing an interesting and challenging problem in Language and Information. To start, you will have to submit a written project proposal. Final projects will be due at the end of the semester. The final presentations will be on December 6th. For your class project, you can elect to write a piece of software or to carry out an experiment and describe it in a research paper. At the discretion of the instructor, other types of projects may be also acceptable.

2 Format

Your proposal should be 2-3 pages long (email is OK). All proposals should include the following parts:

1. Title

2. Introduction (3-4 paragraphs)
   - What is the research problem?
   - Why is it important?
   - Who might be interested in a possible solution to that problem?

3. Overall description (1/2 page + an optional figure or two)
   If writing a piece of software:
   - Where will you get your input data?
   - What is the input? Give an example.
   - What is the output? Give an example.
   - What will the user interface (if any) look like?
   - What processing will be performed?
   If writing an experimental research paper:
   - What is your input data?
   - What hypothesis will you be testing?

4. Method (1 page)
   If writing a piece of software:
• What is the general algorithm that you will be using?
• What programming language will you be using?
• What additional software and/or corpora will you be using?

If writing an experimental research paper:
• Give a step-by-step outline of the data collection process.
• What analytical methods will you be using?
• What software and/or corpora will you be using?

5. Evaluation scheme (1/2 page)
• How will one be able to tell whether you have succeeded?

6. Bibliography
• Pointers to 5–6 papers and/or Web sites that are related to your project.

I will read your proposals and if I have any concerns, I will send you email. If you don’t hear from me by October 31 that would mean that your project proposal is fine and that you can go ahead with it.

3 Final project submission

The following items will be part of your final submission.

1. Slide presentation (for software projects, including demo)
2. Code (if applicable) and data. I have to be able to run your code so please include clear instructions along with any libraries, etc. needed.
3. Full paper or software documentation
4. A Web page with a short project description and the slides from the presentation.

I will ask to see a draft of your Web page two weeks before the due date for the final project. The final Web page will be due at the same time as the project itself. After the end of the class, I will copy all project pages into a shared directory for future reference.

4 Grading

Your project grade will be comprehensive. 75% of the grade will cover the project (including documentation, paper or code, and Web page). 25% will cover the final presentation. Part of your presentation grade will be determined by peer grading – each of you will be judging your peers on five criteria: importance of the project, intellectual value, amount of work done, use of concepts from class, and the presentation itself.

You can get up to 5 bonus points on the project assignment if in addition to the requirements listed above, you produce a nice Web site containing a description of your project, along with any appropriate slides, demo, downloadable code, and/or paper.

Conference track: if you believe that your final project will be a clear contribution to the state of the art in the field, you should talk to me early. If I am convinced that you are on the right track, I will help you turn your results into material for a conference submission. Note that all major conferences in the area (e.g., SIGIR, AAAI, ACL, NAACL-HLT) have deadlines in December of January.

5 Deadline for the proposal

The proposal has to reach me before class time on October 25. The project itself is due on December 5. The final presentations will be in public and will take place on December 6.
6 Sample projects

You are responsible for picking a challenging project. If you are short of ideas, here is a sample list of project topics. Note that some of them may require too much work for a course project so it may be possible to break some projects into smaller pieces.

1. Headline generation
2. Document structure parsing
3. Novelty detection
4. Temporal question answering
5. Fact extraction
6. Question answering using the noisy channel model
7. Word-based network models of the Web
8. Text segmentation
9. Phrase classification
10. Improving multi-document summarization
11. Timeline generation
12. Phrase recognition in IR queries
13. Extracting text from HTML
14. Question answering from semi-structured data
15. Distributional clustering of words and phrases
16. Language reuse and regeneration
17. Word sense disambiguation
18. Trainable summarization
19. Time phrase recognition
20. Multilingual text alignment
21. Query modulation
22. Summarization of speech
23. Text generation using the noisy channel model
24. Identifying paraphrases in text
25. Tracking information provenance
26. Measuring ambiguity
27. Text summarization using the noisy channel model
28. Integrating user models in question answering
29. Meta-evaluation of summarization techniques
30. Determining lifetime of facts
31. Identifying ungrammatical sentences
32. Named entity tagging
33. Building tools for MEAD
34. Building tools for NewsInEssence
35. Building tools for NSIR