

COMS 6998-04, Fall 2012
Statistical NLP for the Web

Project Information

Key Dates :

Sept 19th, 2012 (11:59pm) : Project Proposal Due
Oct 17th, 2012 (11:59pm) : Project Intermediate Report I Due
Nov 21, 2012, Project Intermediate Report II Oral (please make appointments)
Dec 12th, 2012 : (11:59pm) Final Project Due
Dec 12th, 2012 : Final Demo Due
Dec 12th, 2012 : (presentation slot) Final Presentation

Grade Distribution

Project Proposal – 5%
Intermediate Report I – 10%
Intermediate Report II – 20%
Final Report – 30%
Demo – 15%
Final Presentation – 20%

Final Project:

Final project should give you an exciting opportunity to explore a problem of your interest in Natural Language Processing and/or Machine Learning. The research problem you are addressing can be theoretical research question or an applied problem in NLP or some combination of both. Your project cannot be just a programming exercise. The final submission to the project needs to be an end-to-end web or mobile application. You should be able to justify the proposed solution in your final report. The project is meant for you to have a learning experience where you are able to pursue a research topic of your choice in depth and also build an application that can be useful.

Project Proposal [5% of the project grade]

Project proposal should discuss the research problem you are addressing in detail. You should show that you have done some literature review pertinent to the research problem. You should discuss the topics related for the project. Some of these topics could be learning algorithms, data/corpus, annotation/labeling, evaluation method, etc. You can also discuss why your project is original and interesting.

Length: Roughly 2 pages

Intermediate Results I [10% of the project grade]

Report Length

- 1 person team – 4 pages
- 2 person team – 6 pages
- 3 person team – 8 pages

Intermediate Results is to help you guide through the steps of your project and give us an idea about the state of your project. You should clearly state the research question and your proposed solution. You should describe the corpora you are using for the project and the kinds of processing you have done it. If you are using any tools you should describe them as well. You should state the intermediate results you have obtained so far. If the research problem you are exploring makes it difficult to produce intermediate results you should discuss why this is the case in detail. You should also provide the rough plan of how you are going to pursue rest of the project. The overall structure of the report should be formatted to the following structure

- Title
- Abstract/Summary
- Introduction/Motivation
- Related Work
- Data
- System Description
 - NLP/ML Algorithm
 - User Interaction Mockup
 - Experiment
- Results and Discussion
- Conclusion

Please make sure you address these issues

- You need to adhere to the format described above
- You need to provide data description
- You need to have completed basic prototype implementation of the ML algorithm
- You need to have done 1 set of experiments
- You need to present at least 1 table of results
- You need to present a mockup of web interface
- You need to submit through courseworks dropbox

Intermediate Report II (20%)

Grading for Intermediate Report II Oral will be based on

- Demo
 - Is it working?
- Approach
 - Is your approach well thought through
- Results
 - Is the model accuracy real low?
- Discussion
 - Have you thought about ways to improve the model
- Q&A
 - Theory behind algorithms you have used

Final Project Report (30%)

Final project report should be a conference style 8-page paper with overall formatting and division into section as described in Intermediate Report I. The paper template you should use can be downloaded from ACL 2010 website at <http://acl2010.org/authors.html> or http://acl2010.org/authors_sub.html (both latex and word style files available). You can use either latex or word document templates. In your report you should describe the problem, discuss related work, theory of your solution, implementation issues, experiments, results, discussion and critical analysis of your work. I would like to stress the fact that critical analysis of others and your work is essential in any successful research endeavor. Hence, you should think about why your solution is better, why some techniques work while some does not, why your research problem could be interesting to others, how can you improve on your solutions, how does it compare with the most recent related work, etc.

You can take a look at ACL or NIPS website for ideas.

<http://aclweb.org/anthology-new/>
<http://nips.cc/Conferences/>

Report Minimum Length (No Maximum)

- 1 person team – 4 pages
- 2 person team – 6 pages
- 3 person team – 8 pages

Overall formatting of the paper should follow the same requirements as Intermediate Report I restated below (except you should use ACL 2010 style files).

- Title
- Abstract/Summary
- Introduction/Motivation
- Related Work
- Data
- System Description

- NLP/ML Algorithm
- User Interface snapshot
- Experiment

- Results and Discussion
- Conclusion

Please make sure you address these issues

- You need to adhere to the format described above
- You need to provide data description
- You need to have completed the implementation of the ML algorithm
- You need to have done present experimental results
- You need to discuss the results and provide critical analysis of your work
- You need to submit the report through courseworks dropbox

Final Project Demo (15%)

The final project demo will be held on the last day of classes.

Final Project Presentation (20%)

Clarity of the problem description
Originality of the proposed solution
Related Work Knowledge
Description of implementation
Discussions of your experimental results
Critical Analysis of your work