

# Sample Project Description

## CSCI 544 Project

By Dr. Zornitsa Kozareva

### **Class Project**

The purpose of the research project is for the students to learn how to formulate a simple natural language problem/task/application and to experience how to solve it using methods, algorithms and techniques taught in class. The students will conduct experimental evaluation on an interesting dataset and will analyze the obtained results. Students are encouraged to identify new problems/tasks/applications, however we will also provide them with a sample of topics. Building a demo is optional, but will count as bonus points. We strongly encourage the students to work in groups of 2 people.

### **Grading of Course Project**

- Project proposal: (15% of the project grade)
- Mid-term report: (20% of the project grade)
- Final project report: (45% of the project grade)
- Presentation: (20% of the project grade)

### **Project Proposal**

The proposal should try to answer the following questions:

- What is the problem you are solving?
- Why is it interesting and who would use it when solved?
- What data will you use?
- How will you collect and annotate the data?
- What work do you plan to do in the project?
- Which algorithms/techniques/models do you plan to use/develop? Be as specific as you can!
- How will you test your idea?
- How will you measure success?
- What do you expect to submit/accomplish by the end of the semester?

## **Mid-term Report**

By the time of the mid-term report you should complete at least 30% of your project work. The mid-term report is a draft of your final report, but without all experimental results and analysis.

You should submit:

- Introduction to the problem you are working on, why is it important and what makes it challenging to solve
- Review of relevant literature
- Description of data collection process
- Description of data annotation and Kappa agreement scores
- Description of interesting phenomena/statistics in the collected and annotated data
- Description of features used/or planning to be used
- Algorithm description
- Description of challenges you have encountered
- Indicate the parts of the project that have been completed
- Indicate the parts which have not been completed and how do you plan to accomplish them for the final project report

## **Final Project Report**

The final report should follow the ACL format and guidelines (Note the guidelines will be explained in class). During grading we will assess your work based on:

- Introduction/Motivation/ProblemDefinition (10%)
  - What is it the problem you are solving?
  - Why is it interesting?
  - Who would use it when solved? And how would it be used?
  - Why is it challenging?
  - What are the shortcomings and limitations of the existing work?
- Related Work (10%)
  - How does your work relate to those done by others in the NLP field?
  - Provide a citation to the papers you have read, explain briefly what each paper is about, what are the pros and cons of the approach, how

does it compare and contrast to your approach? Did your method improve existing algorithms on the same dataset?

- Data Collection/Annotation (20%)
  - Describe the data you have used in your experimental study
  - Show the number of examples you have used to conduct your study
- Method Description (20%)
  - Provide a detailed description of your algorithm/method, the features used
- Results and findings (30%)
  - Describe how did you evaluate your solution
  - What evaluation metrics did you use?
  - Describe a baseline system
  - How much did your system outperform the baseline?
  - Were there other systems evaluated on the same dataset? How did your system do in comparison to theirs?
  - Show graphs/tables with results
  - Error analysis
  - Suggestions for future improvements
- Style and writing (10%)
  - Writing, grammar, organization and neatness.

### **Project Presentation**

The goal of the presentation is to give you a chance to share with your classmates the exciting problem you have been working on, how you solved it and the interesting findings you discovered. Prepare maximum 8-10 minute presentation, which will be followed by 2 minutes of questions from your instructor, the TA and your classmates.

### **Project Timeline**

- Feb 12: Identifying team members
- Feb 19: Project topics
- Mar 5: Proposal due
- Apr 2: Mid-term report due (data description, preliminary results)
- Apr 23: Final report due

- Apr 23: Project presentations in .ppt or .pptx format due
- Apr 25 – May 2: Project presentations

### **Sample Project**

“*Sentiment Analysis in Twitter*” the goal of the project is to develop an automated machine learning system for sentiment analysis in social media texts such as Twitter. The students should collect experimental data on their own using the twitter API, they will manually annotate the collected tweets, measure the Kappa agreement of their annotations, provide analysis of the data and then focus on building the machine learning classifier(s) using the toolkits already discussed in class. The students will study the effect of different features and feature combinations and will deliver a system capable of identifying the sentiment in the tweets. The project will consist of 2 people.