

# CSCI 6350 Selected Topics in Artificial Intelligence:

## Natural Language Processing (Barbosa S20)

### Graduate Project (Rev1)

#### Purpose

The purpose of the graduate project is to afford students the opportunity to research various modern mechanisms for natural language processing (NLP), conduct comparison and contrast of methodologies, develop software implementing NLP techniques, or conduct research on related topics. In the spirit of academic research, students will present their work in class.

#### Grading

Grading will be made up of the three weighted components described below. Primary references (papers and books vs. web references) published in the past five (5) years must form the nucleus of the research work, and should be used in devising the proposal.

#### I. Project Proposal (20 % of project grade and due on Mar 5, 2020 by 6 PM)

The proposal content should be more than 1 but less than 2 pages long (single spaced, 1 inch margins) and should describe the intent and scope of the project. An additional 1 page references section, with publications related to the subject matter, is required. Topics should be selected from the following:

- Current state-of-the-art analysis – Subjects in this area may include: analysis of existing methods in natural language processing, with focus on the efficiency and effectiveness of the method(s), or on the (in)completeness of common techniques (substantiated by comparative research and analysis).
- Novel methods - Proposed novel or hybridized algorithmic and methodological alternatives to the NLP approaches discussed in class or in the textbook; these methodologies may be implemented as a software model to demonstrate effectiveness.
- Use of word vector embeddings in NLP – Focus on use, manipulation, or enhancements to word vector representations to solve NLP problems.
- New or enhanced applications in one of the following areas:
  - Information extraction
  - Machine translation
  - Sentiment analysis
  - Question answering
- Other projects of sufficient scope and relevance proposed by the student. Such topics must be discussed with the instructor early enough in the topic selection phase to allow for adjustments and modifications prior to the proposal deadline date.

The instructor must approve each proposal, and any programming language, environment, libraries, modules to be used. Python is preferred. Obscure or proprietary languages will not be approved.

## **II. Project Report (50% of project grade – due on Apr 14, 2020 by 6 PM)**

The project report must be a scholarly paper of at least 4 but no longer than 8 single-spaced pages (excluding references) documenting the research and findings, in **PDF form**. The report will have 1 inch margins on the left, right, top, and bottom of each page. The paper should contain graduate-level analysis of the topic and suggested avenues for future research (this precludes a report consisting mainly of figures and source code listings). The report should follow the outline below:

Abstract – A brief introduction (150 - 250 words) to the problem and research to be done.

Introduction – A detailed introduction to the problem being analyzed or implemented, and motivation(s) for why it is an important research area.

Background/Related work – Related works (cited) that inform the project and provide a springboard for its relevance in contributing to the body of knowledge.

Research method – Description of how the research is to be conducted, and the method(s) by which results will be analyzed.

Results and analysis – Description of the actual conduct of the research, presentation of relevant data collected, and analysis of the outcome.

Conclusion and future work – Summary of the project findings, and suggestions for how it may be expanded in the future.

References – Excluded from page count

Appendix (if applicable) – A full description of the coding environment, including libraries and modules used for all projects that involve code. The description should be sufficiently detailed so that the coding environment used can be recreated.

Students are encouraged (though not required) to format the project report in a manner that would enable its quick conversion into a conference or journal paper. In such eventualities, co-authorship with the course instructor will be explored, if sufficient individual contributions warrant. Students retain primary authorship.

Source code – Source code developed for the project report must be submitted in electronic form. The source code should be formatted as a complete project that may be immediately executed by using simple command line invocations, running project *makefiles*, or in an integrated development environment (IDE). Other than short snippets for highlighting or clarifying specific points, source code should not be part of the report.

## **III. Project Presentation (30% of project grade – in class on Apr 14, 16, 21, 23, and 28, 2020).**

An oral presentation on the project, lasting at least 15 minutes but no longer than 17 minutes, must be given in class. A 3 to 5 minute question and answer session will follow each presentation. The presentation should include slides, software demonstrations, and analysis results, as applicable. Note: It is possible that by this point students will have a lot of data to present. Part of the challenge of the presentation is condensing the relevant facts and highlights of the research into 15 to 17 minutes.

**Submit slides 24 hours prior to class start time of the day you will present as PDF (1 slide per page)**

## Submissions

The proposal and project report must be submitted at the beginning of class on the appointed due date. Presentation slides must be submitted 24 hours prior to class time of the day of your presentation. Each of these must be submitted via the handin/Gus system.

Proposal: *handin gpprop project\_proposal.pdf*

Proposal: *handin gprprt project\_report.pdf*

Proposal: *handin gpslides project\_slides.pdf*

## Rubric

The following criteria will be applied in grading each component of the project.

### I. Project Proposal

- Excessive margins: -10%
- Proposal not single-spaced: -10%
- Proposal less than 1 page or greater than 2 pages: -20%
- Proposal late: -100% (must still turn in)
- Proposal content: As graded by instructor

### II. Project Report

- Excessive margins: -10%
- Report not single-spaced: -10%
- Report less than 4 pages or greater than 8 pages: -20%
- Figures and code take up excessive space in report (> 25%): -25%
- Report late: -50%/day (must still turn in)
- Report content: As graded by instructor

### III. Project Presentation

- Presentation less than 15 minutes or greater than 17 minutes: -%5 per 15 seconds off
- Demonstrations don't work or crash: -15%
- Slide presentation relevance, readability, and quality: As graded by instructor
- Presentation clarity and quality: As graded by instructor
- Handling of questions: As graded by instructor