CS-6390 Project, Basic IE System Submission
Due: March 10, 2017 by 11:00pm

Basic System for IE Class Project

Here is a reminder of the criteria for your Basic IE Systems, copied from the original project description:

> *On March 10, each team must submit a Basic IE system for your project and evaluate its performance. Your system should be a functioning IE system and should use a conventional NLP technique to solve your IE task.*

> *The purpose of this project phase is: (1) to make sure that everyone has a working system at this point in the semester, (2) to obtain results for a simple (but non-trivial) approach for solving the problem, and (3) to implement evaluation metrics and assess the performance of the conventional approach.*

> *For most projects, a sensible plan would be to either build a supervised machine learning classifier for the task (if you have sufficient labeled training data) or to implement an algorithm previously used to solve the problem (or a similar problem). The goal should be to create a working system that achieves a respectable level of performance, relative to the difficulty of the task (many IE tasks are quite challenging, so good results will not necessarily be high percentages).*

Basic IE System Submission Requirements

Each team should submit the following for your Basic IE System Submission:

1. **Source code**: please submit all of the source code that *you* wrote (both for the IE task and for the evaluation of your system) and data files or resources that *you* created for your project. You do not need to submit external software packages or data resources (e.g., training/test data) that you used, but you will need to document these in the report.

   Consequently, we will not be running your IE system ourselves, but we may opt to look through the code and/or data that you created in order to better understand how your system works and what you did. If we feel that we need further clarification, we may ask to meet with you and have you demo your IE system for us in the CADE lab, so your code should compile and run there. We don't expect to do this for most teams for the Basic IE System submission, but please be prepared to do so if needed.

2. **Report**: please submit a written report with the following information.

   (1) **Technical Design** (1 page): describe the technical components in your system and explain how they work. Discuss the overall architecture as well as the individual components, and include an in-depth description of the information extraction algorithms that you used. (You don't need to describe the non-IE components in great detail. For example you can just list any data pre-processing steps you may have done.) You *must* also include a list of all

external software packages, tools, and data resources that you used, along with URLs showing where you got them, if appropriate.

(2) **Evaluation Results** (1 page): explain how you evaluated your IE system and present the results of your evaluation. Feel free to include any analyses you may have done to understand the behavior of your system.

(3) **Sample Output**: create an Appendix that shows some sample output of your system. For example, show 2-3 documents (or 2-3 paragraphs from different documents if they are long) and the information that your system extracted from them. The system output may be in the form of BIO labels if you're using a classifier, or XML mark-ups, or extracted strings in a stand-alone template. We just want to see examples of the kind of output that your system is producing. (Don't worry if the output isn't correct – just pick some cases that are fairly typical of your system's behavior.)

In total, your report should be 2 pages in length with an additional Appendix that contains the sample output.

ELECTRONIC SUBMISSION INSTRUCTIONS
(a.k.a. "What to turn in and how to do it")

Only 1 team member needs to submit your project materials! We don't want multiple submissions for a team – that just adds confusion as to which one we should grade.

Please submit an archived (.tar) or zipped (.zip) file that contains the following items on our course's CANVAS site by 11:00pm on Friday, March 10.

1. The source code and data files that you created for your IE system. Please include a README.txt file briefly explaining what each file is.

2. The written report presenting the technical design, evaluation results, and sample output for your Basic IE system. Please submit your report in .pdf format.