

Dear Students;

This is the start of your major assignment for the course. It is a project on one of the topics below. The goal is to investigate the power of NLP on Information Retrieval and Web Search, preferably with Arabic in mind. The first step is to select a topic and write a ONE PAGE Abstract as to the way you understand the project, things you want to achieve and ways to measure success. You have to submit that to the Instructor for approval.

The following conditions apply:

- The topic selected cannot be something you already worked on in the past, cannot be another course project and cannot be your graduation project (no double credit principle).
- You can work in groups of 2 students and in certain cases 3, but the amount of work needs to be proportional to the number of partners in the group. That has to be included in the submitted abstract.
- You need to be able to demonstrate the project to the instructor at the end of the semester.
- You need to submit a full report (one per group) and to present that if needed.

Suggested timeline:

- Team membership and Project selection: Deadline: October 28th, 2017.
- Project and team approval by Instructor: Deadline: October 30th, 2017.
- Discussions and abstract submission and team membership: Deadline: November 8th, 2017.
- Instructor Comments on Project proposal (feedback): November 16, 2017.
- Final Project submission: One week from Last day of Classes.
- Project Discussion: to be announce later

Possible Projects:

- 1- Sentiment Analysis in Arabic: you need to do a classification of sentences, articles, text chunks into positive and negative, maybe with scales. You need to discuss the effects of applying NLP techniques on the process of sentiment analysis. Possible material to analyze: Opinions on products and services, political opinions, textual course evaluations, medical prognoses, and more
- 2- Text similarity: you need to assess the similarity of texts in as far as their meaning is concerned. You need to investigate the effects of NLP on the results. The texts could be in a single language or Cross Lingual: meaning one document is in Arabic and the other is in English.
- 3- Recommender system: an approach using information about the user and his/her previous readings to recommend new reading materials to that individual and study the

effect of NLP on the performance of such a system. The application domain could be health, sports, economics, friends suggestion, professor selection and so on.

- 4- Clustering/classification in a particular domain such as news, health information and so on. Build a model for classifying texts into a set of categories/clusters also employing NLP methods to improve the results. One could also try to use Cross Lingual aspects, meaning clustering both Arabic and English documents into the same category system.

****If you feel strongly about an issue that you would like to work on for this course and meets the requirements, you may discuss that with the instructor as early as possible. It has to be of adequate coverage, related to IR and NLP and uses good quality data and be unrelated to your other course work (including graduation project).**

You will probably need to use NLP tools such as stemmers and Named Entity recognition tools, which should be readily available. You may also need to look for data sets to work with for the training and testing processes. Finding such sets and tools is part of the course requirements and it is fine to get help from your instructor and others. One may also need to use machine learning tools such as Weka or Mallet and it is your responsibility to master the use of such tools as needed for your work. When needed you can consult with people who know such as instructors and students using similar tools. We can talk about that as we go.

Steps

- Pick a topic and form a team and survey related work and submit a proposal abstract: by November 8th, 2017
- Proposal approval by the instructor: by November 12th, 2017
- Work on the project
- Write a report: last week of classes
- Present the project: last week of classes

Evaluation:

Your project will be graded based on the following required components:

- Project proposal (20%)
 - Motivation, problem, and expected outcome of the project.
 - Discussion with instructor prior to deadline is encouraged.
- Project Work (40%)
 - Quality of the results, use of the proper tools, adequacy of testing, ability to introduce changes when needed, quality of test and training data and more.

- Project Presentation: 20%
 - 15 minutes presentation about what you have done for this course project. May include a demo of the system.
- Project report (20%)
 - Detailed written report of your project.
 - Quality requirement is the same as research papers (or graduation project) i.e., in formal written English and rigorous paper format, referencing, sectioning and so on.
 - Due **before** project presentation. I need to check the report before I listen to the presentation. Submit electronically but bring a hard copy to the presentation.

Good luck