Coreference Resolution
Project Assignment

Berfin Aktaş

Due date: September 15, 2018

1 Project Description

The aim of this project is to implement a coreference resolution system, test it on a coreference-annotated dataset and measure the success of the system by the Conll-2012 scoring script.

2 Specification

- Data: **OntoNotes 5.0**: https://catalog.ldc.upenn.edu/ldc2013t19
- Evaluation Method: **Conll-2012 scorer v8.1**: https://github.com/conll/reference-coreference-scorers
- Students are free to integrate third party systems/libraries to their implementations.
- Any programming language can be used.

3 Submission

Expected files (all should be packed in a zip file):

- **README:** Describing the software components of the system (programming language, used libraries, required settings to run the system in different environments etc.). This file should be as informative as possible such that it should contain all the necessary steps to run the system in a different environment.
- **Project Report:** A detailed report around 3000 words presenting the theoretical aspects of the system, such as the implemented clustering/classification algorithms, used learning and testing methods, applied constraints etc. The report should provide the results of the scoring script and also a qualitative discussion on the error types (e.g., what kinds of errors are encountered in the automated system's results, further suggestions to minimize these errors etc.). The report should also include a "related work" section where up-to-date developments in the field are discussed. Reference to five state-of-the-art scientific papers should be sufficient for this section. These references can be chosen from the papers in the course syllabus.
- All project files to run the system (e.g. libraries, configuration files, code files etc.)

3.1 Grading Criteria

- The deadline should be met. (must)
- Software Implementation (%40)
  - The technical/scientific level of the implementation work performed: How complex was the task you did? How much contribution did you make to ready-made components that you could include?
  - The README file: How informative is the README file about the architecture and working principles/settings/requirements of the system?
- Project Report(%60)
  - The report should be submitted as a pdf file. (must)
  - The background description of the research area and previous work: Does it clearly describe the state-of-the-art?
  - The presentation of the work in the report: How readable is the report? How clear is it what you have done in the project?
  - The evaluation of the project: Is the scorer applied correctly? How are the results interpreted?