

CMPS 143 - Assignment 7

Due Friday, March 1st, 11:59 PM

1 Introduction

This homework assignment will expand upon the question answering task introduced in the previous assignment. The new dataset also has new medium difficulty questions for both the old and the new stories. The Input, Output, and Evaluation of your system is the same as for the last assignment and are explained in the Assignment 6 description.

2 The Task

You must improve your question answering (Q/A) system that can process a story and a list of questions, and produce an answer for each question. Your system must conform to the provided input and output specifications. Also, you must **improve your system by using either the constituency parses or dependency graphs (or both)**. Other than these requirements you can implement your system however you want.

The dependency graphs and constituency parses were created using the parsers in Stanford CoreNLP¹. A variety of dependency relation tag sets have been created. Stanford CoreNLP uses the Universal Dependencies² tagset. There is extensive online documentation for Universal Dependencies. One resource³ that you might find particularly useful provides definitions, explanations, and examples of each dependency relation type. Also, Stanford CoreNLP has an online demo at <http://corenlp.run> that displays visualizations of dependency graphs and constituency parses.

Reminder: Note that while we are providing you with some freedom on how you implement your system, this does not mean that you can choose to make only minimum changes to your system from last week. That is, you cannot submit the same system from the last assignment again as your solution to this assignment.

3 Dataset

We have updated the files from HW6 so you MUST download the new dataset for HW7.

The easy questions and the medium questions are designed with different aspects of natural language processing in mind:

- **Easy:** questions that can be answered with simple string matching.
- **Medium:** questions that can be answered with constituency or dependency parses, or advanced regular expressions or string matching on the .story or .sch files.

The kinds of patterns in constituency and dependency parses that the medium questions are based on have been discussed in the class.

¹stanfordnlp.github.io/CoreNLP

²universaldependencies.org/#en

³universaldependencies.org/en/dep/index.html

4 Provided Files

The new dataset is located in the `data` directory. We have also made changes to the `qa_engine` framework so that it is compatible with the new dataset. To get started on this assignment create a new project directory that contains the `data` and `qa_engine` subdirectories. Then copy your Python files from last weeks assignment into the new project directory.

5 What To Turn In

Team submission: Pick one team member to make the team submission. The team submission should be a zip file with the last names of each group member. For example, if the group members were Jane Smith and John Doe then their team submission file would be named `Smith_Doe_6.zip`. The team submission zip file needs to contain the following files:

1. Your question answering system, name it `qa.py`.
2. Your response file that contains the answers to all the questions for all the stories. This should be called `hw7-responses.csv`
3. README including all your team members names and any notes, if necessary, to the grader on how to run your program.
4. Include all files required to run your program other than `qa_engine`.

Individual submission: Each individual should submit the following,

1. Write down who is on your team, and who is responsible for turning in the code.
2. Write a summary of what got done during this phase of the project. This should be represented by a list of tasks and how they were completed.
3. Write down your individual contribution during this phase of the project.
4. Write down the contribution of your other team members for this phase of the project.