

Fall Semester 2017
Iowa State University

ENGL 120 - Language and Computers

Assignment 4

Submission Deadline: 21 OCT 2017, end of the day

Instructions: This assignment consists of two questions and carries a total of 10 marks. Submit your assignment as a *PDF* file and name it as: `your_first_name-your_last_name.pdf` Late submissions are allowed, but will not be awarded full credit.

Question 1

Take the following text snipped from the Wikipedia article on ISU: *"Iowa State University of Science and Technology, often referred to as Iowa State, is a public flagship land-grant and space-grant research university located in Ames, Iowa, United States. It is the largest university in the state of Iowa and the 2nd largest university in the Big 12 athletic conference."* (without the quotes that is).

Go to `corenlp.run` and copy this passage there to see the output of that web application. What are the different outputs it gives? What aspect of language analysis is shown by each output? Think of a possible application scenario for each output (POS, Dependencies, NER, Open IE).

Question 2

What is your idea of a spell checker? How is it developed? If I have a large dictionary of words in a language, will that be sufficient? What are the cases when such a large dictionary is either insufficient or cannot be constructed? (Hint: Think about languages beyond English). What solutions can you think of beyond having a large dictionary for a spell checker? Finally, consider a language like Turkish - a whole sentence can be said in one word (google for examples!). What are the issues one faces with designing a spell checker there?