Assignment 1

Sebastian Ellefsen

January 27, 2020

1 Exercise 1

As described in **Lab1_1.py**, the program read sample.txt and parses each sentence based on it's gender context, i.e is it talking about a male, female, both, or neither. It then calculates the percentage of sentences in each of these categories.

2 Exercise 2

Subtask a and b was combined to a single program that runs based on the command-line arguments.
The program is executed with *$ python3 Lab1_2.py [optional word]*. If given no arguments it runs task A, and if given a word it runs task B with that word.

2.1 A

The index of "sunset" in the word list is found by using [text].index("sunset"). Then to find the whole sentence we start at the found index and move outward until we encounter "." or the start/end of the text. We then have our start and end index.

2.2 B

We fetch the word we want to find from the commandline arguments. We then use the same function as in A to find the whole sentence, but after finding the edges of a sentence we start searching for the next occurence of the wanted word after the end of the previous sentence.

We store the start and end index of each sentence in a list and at the end we join all the words in that range with a space and print it as a sentence.

3 Exercise 3

Skipped because of lack of time

4 Exercise 4

Executed with *$ python3 Lab1_4.py [K] [N]*, where K and N are the same as in the exercise description.

Loops through the first N the words in the brown corpus and converts them to lowercase and counts them with a **Counter()**. Then filters out all words with less than K occurences and prints the remaining words.

5 Exercise 5

Execute with *$ python3 Lab1_5.py*. Fetches an RSS stream of the given page as XML. Then uses the standard python XML parser to read find all posts. Retrieves the title of the posts and splits by " - Søknadsfrist: " which removes the term and we are left with the title and deadline.