20-00-0546-iv Foundations of Language Technology

Homework 7
Categorizing and tagging words (part 1)

17. December 2020

This is a ungraded homework.

7.1 Homework

Homework 7.1 *(6 points) Write code to search the Brown Corpus for particular words and phrases according to tags, to answer the following questions:*[1]

(a) Produce an alphabetically sorted list of the distinct words tagged as MD.
(b) Identify words that can be plural nouns or third person singular verbs (e.g. deals, flies).
(c) Identify three-word prepositional phrases of the form ADP + DET + NOUN (eg. "at the end").
(d) What is the ratio of masculine to feminine pronouns?

Homework 7.2 *(4 points) There are around 230 distinct words in the Brown Corpus having exactly three possible tags if we use the Universal Tagset (the exact number can vary depending on the preprocessing).*

(a) Print a table with the integers 1..10 in one column, and the number of distinct words in the corpus having 1..10 distinct tags in the other column.
(b) For the word(s) with the greatest number of distinct tags, print the sentences from the corpus containing the word, one for each possible tag.

[1] See NLTK-book page chapter 5, exercise 20, page 217