Semantic Computing

Tutorial 2

Summer Semester 2018

Link to comprehensive NLTK tutorial: `http://www.nltk.org/book/`

Exercise 1

Take the following expression and print it in the box notation of Discourse Representation Theory in Python using the *nltk.sem.drt* package. *Anna owns a dog and it bit me.*

Exercise 2

Tokenize and Part-of-Speech (POS) tag the following sentence:
It just tears me apart to see you suffering like that and in tears.
What is the difference in pronunciation, POS tag, and meaning? Use WordNet to retrieve different senses of the word.

Exercise 3

Lemmatize and stem (maybe try different NLTK stemmers) the following words:

- presumably

- provisions

- owed

- abacus[1]

Which stemmer worked better? Which method would you prefer to determine word frequency information of a text corpus?

Exercise 4

Load the words of Macbeth *gutenberg.words("shakespeare − macbeth.txt")* from the gutenberg corpus *from nltk.corpus import gutenberg*. Perform the following operations:

a) Print out the length of all the words in the text.

b) Lowercase all words and remove puncutation and numbers.

c) Lemmatize the result of b), keep only unique lemmas, and print the result.

d) Stem the results of b), keep only unique stems, and print the result.

e) Compare the count of a)-d). What can you observe?

f) Create a list of words that are in WordNet from the lemma list. What kind of words are kept and which ones not?

---

[1] A square object with small balls on wires, used for counting