

19CSE453 Natural Language Processing Assignment 2

Accessing Text Corpora and Lexical Resources

1. Gutenberg corpus :
 - a. List the different texts in the corpus
 - b. Print the name of any text (e.g: text2) and find the length of it .
 - c. Find the words in this text in sorted order.
 - d. Find the length of this set of words
 - e. Find the concordance of a particular word in that text.(e.g: affection in text2)
 - f. Find the words similar to the given word
 - g. Find the file identifiers in this corpus
2. Brown corpus
 - a. Print the different categories/genre in this corpus .
 - b. Find the conditional frequency distribution of the modal verbs: ['can' , 'may','could','might',' must', 'will'] in each genre.
3. Wordnet
 - a. Find the synsets of a given word (e.g: motor car)
 - b. Find the name and definition of each .
 - c. The different synsets of a word (with many meanings eg:car)
 - d. Find the hypernyms and meronyms