EDAN20
Harald Nordgren
900404-1936

Lab Report 2

We downloaded training and test sets, tagged some sentences with *<ng>* and *</ng>* to delimit noun groups. And experimented with *brackets_to_iob2.pl*.

For *tag_complete_file* we had an original accuracy of 59.04% (precision: 69.20%, recall: 51.48%) with *train.txt* for noun groups (NP).

By adding tags NNS, CD, FW, $, NNPS and PRP$ (corresponding to common nouns plural, numerals, foreign words, dollar, proper nouns plural and possessive pronouns, respectively) we managed to raise our harmonic mean to 75.39% (precision: 74.43%, recall: 76.37%). The test set then gave a similar accuracy.

We then did the same for verb groups, adding *verb_group([V]) --> verb(V)* to *ng.pl* and the tags VB, VBD, VBG, VBN, VBZ and VBP. However, now we only reached an accuracy around 50%.