

TP3

Natural Language Processing

HMM

In this experiment we implement an unigram and a trigram hidden Markov models to identify gene names in biological text.

The zip file with the results contains :

- the original files for *countfreqs.py*, *evaltagger.py*, *gene.train*, *gene.dev*, *gene.key*, *gene.test*
- the *gene.counts* file that is based on the train file and is used for computing emission and transition probabilities
- the *gene.dev.p1.out* and *gene.dev.p2.out* that contain tagging results from the unigram and trigram taggers applied on the development set
- the *gene.test.p1.out* and *gene.test.p2.out* that contain tagging results for the test file from the unigram and trigram taggers
- the *tp3Trigram.py* and *tp3Unigram.py* that contain the source codes for the taggers