UvA / Natuurlijke Taalmodellen en Interfaces / 2015-2016
Deel B step 1

Yorick de Boer [10786015]

February 2016

1 Introduction

The subject of this assignment was creating a binarized version of trees written as S-Expressions. To test whether the binarization was succesfull a pre-compiled de-binarization program in Java was given.

2 Method

A fragment in the original treebank (i.e. a sub-tree) looks like:

```
(NP (NNP Rolls-Royce) (NNP Motor) (NNPS Cars) (NNP Inc.))
```

After binarization, it should look like:

```
(NP (NNP Rolls-Royce) (@NP-> NNP (NNP Motor) (@NP-> NNP NNP (NNPS
Cars) (@NP-> NNP NNP NNPS (NNP Inc.)))))
```

The binarization program was written in Python 3. It can be executed by running the following command:

```
python3.5 b1step1.py -infile [input] -outfile [output]
```

Debugging can be done with:

```
python3.5 b1step1.py -debug
```

3 Results

By running the supplied de-binarizer program an average F1 score of 76.59 was calculated. This is in line with the given F1-score in the assignment description.