University of Trento
Department of Information Engineering and Computer Science

Language Understanding System

first project

Roberto Zen

April 16, 2015

Contents

1 Outline

The report is structured as follows. Section describes ... Section 2 describes how I used the *FST* and *GRM* tools for training and testing sequence labeling and it shows the results of applying these tools. Section 3 describes the same as section 2 but using the *CRF++* tool. Section 4 shows how text classification is made using *Naive Bayes*. The last section states the results and the conclusion of the conducted work.

2 Data Analysis

The given data set is composed as follows. Table 1 shows more details about the given files.

File name	Used for	Word count	Token count
NLSPARQL.test.feats.txt	FST	0	0
NLSPARQL.train.feats.txt	FST	0	0
NLSPARQL.test.data	CFF++	0	0
NLSPARQL.train.data	CFF++	0	0
NLSPARQL.train.tok	Naive Bayes	0	0
NLSPARQL.train.utt.labels.txt	Naive Bayes	0	0
NLSPARQL.test.tok	Naive Bayes	0	0
NLSPARQL.test.utt.labels.txt	Naive Bayes	0	0

Table 1: Details of the dataset.

3 Evaluation

4 Sequence Labeling in CRF++

5 Feature sets

6 Tool configuration

7 Conclusion