

Slovene NLTK Tagger

Niko Colnerič

Faculty of Computer and
Information Science
University of Ljubljana

Nejc Banič

Faculty of Computer and
Information Science
University of Ljubljana

December 22, 2011

Abstract

TODO - NIKICC

1 Introduction

TODO - NIKICC

2 JOS Corpus

TODO - BANIČ

3 NLTK Trainer

TODO - NIKICC

4 Taggers description

4.1 Trigram tagger

TODO - NIKICC

4.2 Brill tagger

TODO - NIKICC

4.3 Naive Bayes tagger

TODO - BANIČ

5 Implementation

5.1 Corpus transformation

TODO - BANIČ

5.2 Training the tagger

TODO - NIKICC

5.3 Usage with NLTK

TODO - BANIČ

5.4 Encoding problems

TODO - BANIČ

6 Conclusion

TODO - BANIČ

References

[1] <http://nl.ijs.si/jos/>