

Bachelorarbeit NLP-Plattform: Integration und
Evaluation von POS-Tagging-Algorithmen

Zusammenfassung

Im Themenbereich *Natural Language Processing* (kurz NLP) versucht die Informatik, natürliche Sprachen für Algorithmen zugänglich und interpretierbar zu machen. Ein wichtiger Teil von NLP ist die Identifizierung der syntaktischen Bedeutung von Wörtern (*Parts of Speech*, kurz POS) in gesprochener Sprache, und deren Zuweisung in Form von Tags (POS-Tagging). Für diese Aufgabe existiert eine Vielzahl unterschiedlich leistungsfähiger und robuster Algorithmen.

Im Rahmen dieser Arbeit wurde eine Sammlung solcher POS-Tagging-Algorithmen zusammen mit einem Evaluationssystem als Operatoren im Programm RapidMiner (zur Verfügung gestellt von RapidMiner GmbH) implementiert.

Zuerst untersucht diese Arbeit das Konzept und bestehende Lösungen von POS-Tagging. Dann wird formalisiert, welche Problem- und Aufgabenstellungen in der zu implementierenden Erweiterung auftreten. Daraufhin wird die Implementierung im Detail beschrieben und abschließend wird ein Testlauf des Evaluationsrahmenwerks für die POS-Tagger durchgeführt.

Bachelor Thesis NLP-Plattform: Integration and
Evaluation of POS-Tagging-Algorithms

Summary

The field of Computer Science called *Natural Language Processing* (NLP) attempts to make natural language accessible for machines and algorithms. One of the major parts of NLP is the identification of the syntactic roles of words and symbols (*Parts of Speech*, POS in short) in spoken language, and tagging them as such POS (POS-Tagging). There is a multitude of differently capable and robust algorithms for this task.

In this project, such POS-Tagging algorithms and a system to evaluate them is implemented in form of operators in an extension to the program RapidMiner (by RapidMiner GmbH).

First, the Thesis explores the Concept and already existing Solutions in the Field of POS-Tagging. Then, emerging problems and tasks that appear during implementation of this extension are formalized. After that, the actual implementation is described in detail, and a test of the implemented evaluation methods for POS-Taggers is executed.