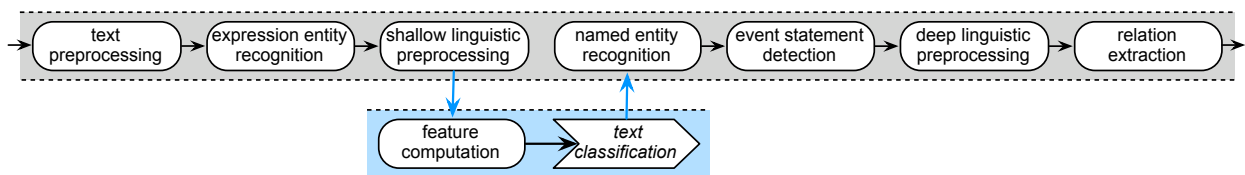


Masterarbeit

Effizienz und Effektivität der Integration von Textklassifikation in Information-Extraction-Pipelines

Motivation

Innerhalb von Information-Mining-Anwendungen dienen Textklassifikatoren der Filterung relevanter Texte und Information-Extraction-Pipelines der Analyse von Texten. Üblicherweise werden Filterung und Analyse als getrennte Phasen umgesetzt, obwohl die meisten Textklassifikatoren auf Ergebnissen von Information-Extraction-Algorithmen basieren. Die so entstehende Redundanz lässt sich durch Integration der Textklassifikation in die Information-Extraction-Pipeline vermeiden. Je früher dabei Texte gefiltert werden, desto weniger Analysen sind auf irrelevanten Texten auszuführen, aber je mehr Analysen auf einem Text ausgeführt werden, desto mehr Informationen stehen für die Filterung zur Verfügung. Dies impliziert einen Trade-off zwischen Effizienz und Effektivität der umgebenden Anwendung.



Aufgabenstellung

Ziel der Masterarbeit ist es, den Trade-off zwischen Effizienz und Effektivität anhand einer gegebenen Information-Extraction-Pipeline mit Algorithmen A_1, \dots, A_k und einer Menge alternativer Textklassifikatoren y_1, \dots, y_k konzeptuell und experimentell zu untersuchen, wobei jeder Klassifikator y_i die Ergebnisse von Algorithmus A_1, \dots, A_i benötigt. Dafür soll die Integration von Textklassifikation in Information-Extraction-Pipelines geeignet modelliert und analysiert werden. Eine experimentelle Evaluierung findet auf Basis einer geeigneten Auswahl bestehender Textdokumente und Algorithmen statt, wofür passende Textklassifikatoren zu trainieren und in Java innerhalb des Frameworks Apache UIMA einzubinden sind. Darauf aufbauend soll betrachtet werden, welche Auswirkung die Integration von Textklassifikation für das am s-lab entwickelten Verfahren zur Laufzeitoptimierung von Information-Extraction-Pipelines hat.

Voraussetzungen

Vorwissen im Bereich der Information Extraction und des maschinellen Lernens

Bearbeiter
Steffen Beringer

Wissenschaftliche Leitung
Prof. Dr. G. Engels

Ansprechpartner

s-lab – Universität Paderborn
Henning Wachsmuth
Raum E4.107
Tel: +49 (0) 5251 / 60-3359
Email: hwachsmuth@s-lab.upb.de