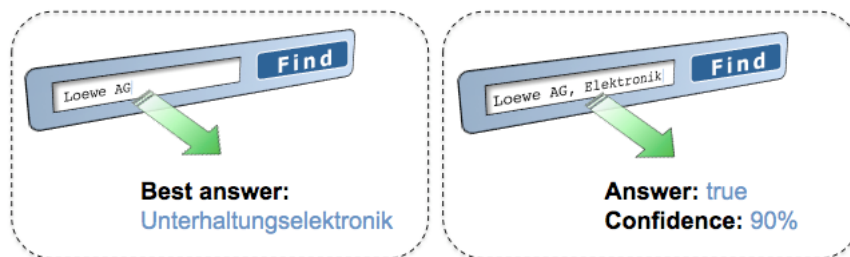


## Bachelorarbeit

# Webbasiertes Mapping zwischen Firmen und Branchen

### Motivation

In einem vom Bundesministerium für Bildung und Forschung geförderten Projekt wird aktuell an der automatischen Erkennung und Verarbeitung deutschsprachiger Aussagen über die Umsätze von Firmen und Branchen aus textuellen Dokumenten gearbeitet. Darauf aufbauend sollen aggregierte Aussagen über Branchen abgeleitet werden, so dass die Zugehörigkeit von Firmen zu Branchen zur Laufzeit überprüft werden muss.



### Aufgabenstellung

Ziel der Arbeit ist es, im Umfeld des Projekts webbasierte Verfahren zu konzipieren und in einem Java-Programm umzusetzen, mittels derer sich die Branche zu einer Firma bzw. die Zugehörigkeit einer Firma zu einer vom Nutzer spezifizierten Branche finden lässt. Welche Mechanismen dafür eingesetzt werden (z.B. statistische Verfahren, Wortfeldanalysen, ...), ist geeignet zu entscheiden. Beispiele für mögliche Funktionalitäten zeigen die beiden Abbildungen.

Die Klassifikation soll so erfolgen, dass sie für viele Firmen hintereinander ausgeführt werden kann und dabei kaum spürbare Wartezeit erfordert. Dabei kann die Granularität der Branche variieren (etwa Elektronik vs. Unterhaltungselektronik). Die Güte der implementierten Verfahren soll anhand der Ergebnisse evaluiert werden.

### Nützliches Vorwissen

Umgang mit Web-Anfragen in Java, Techniken aus dem Bereich des Information Retrieval

### Bearbeiter

Natascha Pranjic

### 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