



UNIVERSITÄT  
LEIPZIG

UNIVERSITÄT LEIPZIG

ABTEILUNG AUTOMATISCHE SPRACHVERARBEITUNG

FORTGESCHRITTENE METHODEN DES INFORMATION RETRIEVAL

## Bericht zum Laborprojekt

Jeremy Puchta – Jonathan Lange – Ali Al-Ali

5. März 2019

## **Inhaltsverzeichnis**

|          |                                     |          |
|----------|-------------------------------------|----------|
| <b>1</b> | <b>Einleitung</b>                   | <b>1</b> |
| <b>2</b> | <b>Datenakquise und -analyse</b>    | <b>1</b> |
| <b>3</b> | <b>Architektur</b>                  | <b>2</b> |
| <b>4</b> | <b>Evaluation</b>                   | <b>3</b> |
| <b>5</b> | <b>Zusammenfassung und Ausblick</b> | <b>4</b> |

# 1 Einleitung

Der vorliegende Bericht erläutert den Aufbau sowie die Funktionsweise der Suchmaschine **Historical News Search**, welche im Rahmen des Laborprojektes innerhalb des Moduls *Fortgeschrittene Methoden des Information Retrieval* im Wintersemester 2018 / 2019 erstellt wurde. Ziel des Laborprojektes war es, die vermittelten Vorlesungsinhalte zu vertiefen und diese bei der Erstellung einer domänenspezifischen Suchmaschine umzusetzen. Die vorgestellte Suchmaschine dient der Exploration von historischen Nachrichten. Verschiedene Szenarien sind als Anwendungsfälle für eine solche Suchmaschine denkbar. Ein Nutzer kann nach der Berichterstattung zu Personen, Gruppen, Orten oder bestimmten historischen Ereignissen suchen. Als Beispiel für Suchanfragen können hierfür `reichstagswahl 1930`, `max schmeling` oder `nsdap` genannt werden. Eine weitere Möglichkeit der Nutzung stellt die Ahnenforschung dar, bei welcher ein Nutzer in Zeitungsberichten nach Informationen über seine Vorfahren suchen kann und somit gegebenenfalls mehr über seine Familienhistorie erfährt.

Im folgenden Kapitel wird zunächst näher auf den Datensatz eingegangen und mit Methoden der deskriptiven Statistik analysiert. Die Architektur und der Aufbau der Suchmaschine werden in Kapitel 3 näher dargestellt. Dazu zählt insbesondere die Erstellung des Indizes sowie die Umsetzung des Suchprozesses über alle beteiligten Komponenten des Systems. In Kapitel 4 werden zunächst die Evaluationsmethodik sowie die Resultate der Evaluation erläutert. Weiterhin wird dargestellt, welche Anpassungen vorgenommen wurden, um die Effektivität der Suchmaschine zu verbessern. Abschließend erfolgt in Kapitel 5 eine Zusammenfassung des Projektes und es wird ein Ausblick in Bezug auf eine Weiterentwicklung der Suchmaschine geliefert.

## 2 Datenakquise und -analyse

### **3 Architektur**

## 4 Evaluation

## **5 Zusammenfassung und Ausblick**

## Literatur