



TECHNISCHE HOCHSCHULE NÜRNBERG  
GEORG SIMON OHM

Fakultät Elektrotechnik Feinwerktechnik Informationstechnik

# Entwicklung eines Suchalgorithmusprototypen zur Bewertung von Suchergebnissen verschiedener Kategorien

Studienarbeit im Studiengang Software Engineering

vorgelegt von

Marc Jonas Roser

Matrikelnummer 364 7316

Betreuer:

Prof. Dr. Hans-Georg Hopf

© 2022

Dieses Werk einschließlich seiner Teile ist **urheberrechtlich geschützt**. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Autors unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen sowie die Einspeicherung und Verarbeitung in elektronischen Systemen.

## **Kurzdarstellung**

Das Ziel der vorliegenden Studienarbeit ist es, eine bestehende Datenbank mit multimedialen Inhalten möglichst effizient nach unterschiedlichen Kriterien zu durchsuchen. Suchergebnisse sollen nach bestimmten Kriterien gewichtet, gefiltert und sortiert werden. Vorschläge für eine weiterführende Navigation auf der Suchergebnisseite sollen angeboten werden, Suchergebnisse sollen dazu nach Kontext und Wahrscheinlichkeiten gewichtet angezeigt werden. Das theoretische Fundament dieser Arbeit stellt die wissenschaftliche Betrachtung der Methoden zur Bewertung der Relevanz von Suchergebnissen dar. Die Arbeit untersucht die Möglichkeit, einen Suchbegriff so zu analysieren, dass ein Nutzer die bestmögliche Ergebnisliste bzw. zielgerichtete weiterführende Navigationsmöglichkeiten erhält. Die bestehende Anwendung „Crossload“ wird vorgestellt, um dem Leser einen Kontext zu bieten, in der sich die Entwicklung bewegt.

## **Abstract**

The goal of this student research project is to search an existing database with multimedia content as efficiently as possible according to various criteria. Search results are to be weighted, filtered, and sorted according to certain criteria. Suggestions for further navigation on the search results page are to be offered, and search results are to be displayed weighted according to context and probabilities. The theoretical foundation of this work is the scientific consideration of methods for evaluating the relevance of search results. The work examines the possibility of analyzing a search term in such a way that a user receives the best possible list of results or targeted further navigation options. The existing application "Crossload" is presented to provide the reader with a context in which the development takes place.

## Eidesstattliche Erklärung

Hiermit versichere ich, Marc Jonas Roser, ehrenwörtlich, dass ich die vorliegende Studienarbeit mit dem Titel: „Entwicklung eines Prototypen eines Suchalgorithmus zur Bewertung von Suchergebnissen verschiedener Kategorien“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden. Ich versichere zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

Nürnberg, 25.09.2022

Marc Jonas Roser



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Relevanz des Themas	1
1.2	Ausgangssituation	1
1.3	Zielsetzung	2
<b>2</b>	<b>Grundlagen</b>	<b>3</b>
2.1	Methoden zur Bewertung der Relevanz von Suchergebnissen	3
2.2	Bisherige Suchfunktion	3
2.3	Crossload	3
<b>3</b>	<b>Anforderungen und Problemanalyse</b>	<b>5</b>
3.1	Vorgehensweise	5
3.2	Funktionale Anforderungen	5
3.3	Nichtfunktionale Anforderungen	5
<b>4</b>	<b>Konzeption</b>	<b>7</b>
4.1	Analyse der Arbeitspakete	7
4.2	Konzeption	7
<b>5</b>	<b>Entwicklung des Prototyps</b>	<b>9</b>
<b>6</b>	<b>Evaluation</b>	<b>11</b>
<b>7</b>	<b>Diskussion der Ergebnisse</b>	<b>13</b>
7.1	Abgleich der Ergebnisse mit den gestellten Anforderungen	13
7.1.1	Funktionale Anforderungen	13
7.1.2	Nichtfunktionale Anforderungen	13
7.2	Identifizierte Probleme	13
7.3	Bewertung	13
<b>8</b>	<b>Ausblick</b>	<b>15</b>
	<b>Abbildungsverzeichnis</b>	<b>17</b>
	<b>Tabellenverzeichnis</b>	<b>19</b>

**Literaturverzeichnis . . . . . 21**

**Glossar . . . . . 23**

# Kapitel 1

## Einleitung

### 1.1 Relevanz des Themas

Suchalgorithmen und relevante Suchergebnisse sind derzeit so relevant wie noch nie. Dabei wollen die Benutzer einer Suchmaschine in Sekundenbruchteilen Ergebnisse, die am besten zu ihrem Suchbegriff passen, ohne sich dabei viel Gedanken über die Formulierung eines solchen Begriffes zu machen. Ein Beispiel für einen solchen Algorithmus ist Google, welches seit den frühen 2000ern einen kometenhaften Aufstieg in der Welt der Suchmaschinen hinter sich hat, was anhand der erreichten Werbeeinnahmen sichtbar wird.<sup>[1]</sup> Google ist im Vergleich zu anderen Suchmaschinen so stark verbreitet,<sup>[2]</sup> dass mittlerweile sogar der Duden das Verb „googeln“ als eigenen Begriff für die Recherche im Internet führt.<sup>[3]</sup> Dabei stellt sich für die Entwicklung eigener Produkte die Frage, wie aus einem Suchbegriff, der meist nur aus wenigen Wörtern bis zu einem ganzen Satz besteht, relevante Suchergebnisse gefunden werden können. Dies würde zur Akzeptanz der Nutzer im Hinblick auf die entwickelte Funktionalität führen, da gewünschte Ergebnisse schneller und ohne großen Aufwand gefunden werden können.

### 1.2 Ausgangssituation

Derzeit besteht bei Crossload4, einer Plattform zum Durchsuchen und Anhören einer umfassenden Predigt Datenbank, eine Datenbank mit einer Such API auf Basis von Spring Boot und SOLR. Diese teilt auf der Suchergebnisseite die Ergebnisse nach Kategorien auf und somit können nur schwer übergreifende Suchanfragen getätigt werden. Zwar werden alle Treffer auf der gleichen Seite angezeigt, doch durch die Aufteilung nach Kategorien werden Ergebnisse gewisser Kategorien über anderen gezeigt, auch wenn niedrig positionierte Kategorien relevantere Ergebnisse enthalten.

### 1.3 Zielsetzung

Das Ziel der vorliegenden Studienarbeit ist es durch eine theoretische Betrachtung der Bewertung der Relevanz von Suchergebnissen und der anschließenden Entwicklung eines Prototypen, ein bestehendes Produkt zu erweitern. Diese Erweiterung umfasst, die nach Kontext und Wahrscheinlichkeiten gewichtete und gefilterte Suche über eine Datenbank mit Datentypen verschiedener Kategorien bei der zusätzlich Vorschläge zur weiteren Navigation auf der Suchergebnisseite gegeben werden sollen.



## Kapitel 2

### Grundlagen

#### 2.1 Methoden zur Bewertung der Relevanz von Suchergebnissen

#### 2.2 Bisherige Suchfunktion

#### 2.3 Crossload



## **Kapitel 3**

### **Anforderungen und Problemanalyse**

#### **3.1 Vorgehensweise**

#### **3.2 Funktionale Anforderungen**

#### **3.3 Nichtfunktionale Anforderungen**



## Kapitel 4

### Konzeption

#### 4.1 Analyse der Arbeitspakete

#### 4.2 Konzeption



## Kapitel 5

### Entwicklung des Prototyps





## Kapitel 6

### Evaluation



# Kapitel 7

## Diskussion der Ergebnisse

### 7.1 Abgleich der Ergebnisse mit den gestellten Anforderungen

#### 7.1.1 Funktionale Anforderungen

#### 7.1.2 Nichtfunktionale Anforderungen

### 7.2 Identifizierte Probleme

### 7.3 Bewertung



## Kapitel 8

### Ausblick



## Abbildungsverzeichnis





## **Tabellenverzeichnis**



## Literaturverzeichnis

- [1] Alphabet, “Werbeumsätze der google-websites weltweit in den jahren 2001 bis 2021,” 2022.
- [2] NetMarketShare, “Marktanteile der suchmaschinen - mobil und stationär 2022,” 2022.
- [3] Duden, “Wörterbuch "googlen",” 2022.



# Glossar

**API** Application Programming Interface. I, 1

**Crossload** Plattform zum Durchsuchen und Anhören einer umfassenden Predigt Datenbank. I

**SOLR** Open Source Suchframework der Apache Foundation. I, 1