



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

Glossar

API

Application Programming Interface.

Crossload

Plattform zum Durchsuchen und Anhören einer umfassenden Predigt Datenbank.

Lucene oder Apache Lucene

Open Source Suchbibliothek der Apache Foundation..

SOLR oder Apache SOLR

Open Source Suchframework der Apache Foundation.

Suchmaschine

Eine Anwendung, die gezielt Ergebnisse aus dem Internet für den Nutzer aufbereitet und sortiert. Englisch: „Search Engine“.

Trial-and-Error

Wiederholtes Ausprobieren ohne Erfolgsgarantie mit Änderung der Startparameter um zum gewünschten Ziel zu gelangen..

Inhaltsverzeichnis

1. Einleitung	1
1.1. Relevanz des Themas	1
1.2. Ausgangssituation	1
1.3. Zielsetzung	1
2. Theoretische Grundlagen	3
2.1. Relevanz	3
2.2. Methoden zur Bewertung von Relevanz	3
2.2.1. Textuelle Relevanz	4
2.2.2. Relevanz durch Attribute	4
2.2.3. Hyperlink Relevanz	4
2.2.4. Relevanz durch Nutzerverhalten	5
2.2.5. Performance	5
2.3. Auswertung der Relevanz von Suchergebnissen	5
2.4. SOLR	6
3. Anforderungen und Problemanalyse	7
3.1. Vorgehensweise	7
3.2. User Stories	7
4. Konzeption	9
4.1. Analyse der Arbeitspakete	9
4.2. Konzeption	9
5. Entwicklung des Prototyps	10
6. Evaluation	11
7. Diskussion der Ergebnisse	12
7.1. Abgleich der Ergebnisse mit den gestellten Anforderungen	12
7.2. Identifizierte Probleme	12
7.3. Bewertung	12
8. Ausblick	13
Anhang	A
I. Bilder	A
Abbildungsverzeichnis	C
Tabellenverzeichnis	D

Literaturverzeichnis E

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.¹ Google ist im Vergleich zu anderen Suchmaschinen so stark verbreitet², dass mittlerweile sogar der Duden das Verb „googeln“ als eigenen Begriff für die Recherche im Internet führt.³ 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 zur Durchsuchung einer umfangreichen Pre-digt Datenbank, welche mit einer Such API auf Basis von Spring Boot und SOLR oder Apache SOLR ausgestattet ist. 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.⁵

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 Prototyps, ein bestehendes Produkt zu erweitern. Diese Erweiterung umfasst, die nach Kontext

¹Siehe 1

²Siehe 2

³Vgl. Duden [1]

⁴Siehe Crossload.org [2]

⁵Siehe 3

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.

2. Theoretische Grundlagen

2.1. Relevanz

Relevanz ist allgemein beschrieben eine Beziehung zwischen einem Individuum, dem zeitlichen Rahmen, in welchem dieses eine Information benötigt und einer beliebigen Information.¹ Das bedeutet, dass Relevanz von Person zu Person unterschiedlich ist, da zum einen diverse Informationen nur zu einer bestimmten Zeit notwendig bzw. wichtig sind und der Kontext der benötigten Information sich ständig ändert.

Um zu verstehen, woher die Relevanz stammt bzw. in der Informationstechnik verwendet wird, ist es wichtig die Gewinnung von Informationen aus Objekten (*Information Retrieval*) zu verstehen. Dieser Teil der Wissenschaft beschäftigt sich mit dem präzisen Abruf von Informationen, um den das Informationsbedürfnis (*Information need*) eines Nutzers zu stillen. Das Informationsbedürfnis wird hierbei von einem idealen Inhalt gestillt, stellt also die Spezifikation eines idealen Inhalts dar. Diese Spezifikation geht aber über den reinen textuellen Inhalt der Suche hinaus. Die Relevanz ist dabei die Aktivität bzw. Praxis um diesen idealen Inhalt zu finden.²

2.2. Methoden zur Bewertung von Relevanz

Eine Suchmaschine gibt nach Anfrage Webseiten sortiert nach der Relevanz der Ergebnisse abhängig zum gegebenen Suchbegriff des Nutzers. Die Schwierigkeit dabei ist die Bestimmung der Relevanz für eine beliebige Website. Die dafür genutzten Funktionen und Methoden werden allerdings von den Unternehmen geheim gehalten, um einen Missbrauch ihrer Suchmaschine zu verhindern. Dennoch sind die am häufigsten genutzten Merkmale bekannt und in einigen wissenschaftlichen Arbeiten untersucht worden.³ Da Moderne Suchmaschinen nutzen dutzende oder gar hunderte verschiedener Methoden um Features um die Relevanz der verfügbaren Suchergebnisse zu bewerten, wird im folgenden nur auf einige bekannte Methoden eingegangen.

Zur Einfachheit wird von der Webapplikation Crossload abstrahiert und stattdessen Beispiele aus der Internetsuche verwendet, welche zum Beispiel mit Google, Bing, Ecosia oder anderen Suchmaschinen üblich ist.

¹Vgl. Bookstein S. 1 [3]

²Vgl. Manning, Raghavan, Schütze [4]

³Vgl. Zaragoza, Najork, S. 1 [5]

2.2.1. Textuelle Relevanz

Das einfachste Merkmal für die Bewertung der Relevanz ist den kompletten Inhalt nach der textuellen Relevanz zu bewerten. Da natürliche Sprache, die meist für Suchergebnisse genutzt wird, generell ungenau ist, wird mit sogenannten „Matching Functions“ versucht auch ungefähre Übereinstimmungen in einem Fließtext zu finden. Einige der verwendeten Funktionen um die textuelle Relevanz zu bewerten sind dabei:⁴

- Die Anzahl der Treffer für den Suchterm oder Abwandlungen
- Position des Suchterms (früheres Vorkommen)
- Seiten Struktur (für Webseite: Ist der Term eine Überschrift o. ä.)
- Grafisches Layout (für Webseiten: Ist der Term z. B. farblich markiert)

2.2.2. Relevanz durch Attribute

Des Weiteren ist es auch möglich den durchsuchten Objekten Attribute zuzuweisen, um für Schlagwörter relevantere Ergebnisse zu erlangen. Diese können entweder von Nutzern selbst bestimmt werden, wie z. B. bei der Website Flickr⁵, um Bilder für bestimmte Themen höher werten zu lassen oder werden von Algorithmen aufgrund von Bilderkennung automatisch zugewiesen.

Ein Beispiel hierfür ist Google, welches eine frei verwendbare Machine Learning API⁶ oder eine direkte Integration, in die Google Fotos App anbietet, welche die gemachten Bilder automatisch in verschiedene Kategorien aufteilt.⁷

2.2.3. Hyperlink Relevanz

Für Suchergebnisse im Internet oder andere miteinander verlinkte Seiten, wie z. B. in internen Dokumentationsseiten, Wikis o. ä., können auch die Hyperlinks, die auf eine andere Seite verlinken genutzt werden die Relevanz eines Ergebnisses zu bestimmen. Ein Hyperlink besteht hierbei aus dem angezeigten Text auf der Quellseite und einem Link zur Zielseite oder auf einen bestimmten Abschnitt dergleichen. Dies ist aber kein automatischer Prozess, sondern jeder Link wird von Menschen gesetzt. Aus diesem Grund kann man hier von „menschlicher Intelligenz“ sprechen.⁸

Um einen Treffer höher zu gewichten, ist eine Option die Anzahl an Verlinkungen auf eine Seite zu zählen und absteigend zu sortieren.⁹ Alternativ kann der angezeigte Linktext noch

⁴Vgl. Zaragoza, Najork, S. 1 [5]

⁵Vgl. Liu et.al., S. 1-3 [6]

⁶Vgl. Google ML Dokumentation [7]

⁷Vgl. Google Fotos [8]

⁸Vgl. Zaragoza, Najork, S. 2 [5]

⁹Vgl. Marchiori [9]

zusätzlich als eine Art Attribut (2.2.2) oder erweiterte textuelle Referenz (2.2.1) gesehen werden, der dann bei der Auswertung einer Suche mitverwendet wird.¹⁰

2.2.4. Relevanz durch Nutzerverhalten

Um unabhängiger von manuellem Verlinken zwischen Seiten zu werden, haben bekannte Suchmaschinen auch Möglichkeiten entwickelt, die Anzahl der „erfolgreich“ gefundenen Treffer zu zählen und als relevanter zu gewichten. Im Umfeld einer Internetsuche wäre der „erfolgreich“ gefundene Treffer ein Klick auf die entsprechende Website. Diese können entweder live oder durch Auswertung von Log Dateien analysiert werden. Andere Wege um die Anzahl an Besuchen auf einer Website zu messen, umfassen Tracking Methoden, Toolbars oder Werbung. Diese Methode ist überaus erfolgreich, da hier von einer Art Schwarmintelligenz ausgegangen wird, die Nutzern für die gleiche Suche Ergebnisse anzeigt, die schon viele Benutzer davor angeklickt haben.¹¹

2.2.5. Performance

Da Suchmaschinen dem Nutzer eine bestmögliche Benutzererfahrung, auch bekannt als User Experience, ermöglichen wollen, sollen die gefundenen Webseiten dies bieten. Eine Möglichkeit dies zu messen ist die Performance einer Website. Dies umfasst die Ladegeschwindigkeit, Speicherverbrauch und benötigte Leistung um die Seite komplett anzuzeigen. Da dies nicht für x-Millionen Treffer bei jeder Suchanfrage getestet werden kann, werden mögliche Suchtreffer vorher indiziert und nach der Performance untersucht. Dadurch entsteht ein Performance-Score, welcher dann für die Relevanz verwendet werden kann.¹²

2.3. Auswertung der Relevanz von Suchergebnissen

Um letztendlich Ergebnisse mit der höchsten Relevanz zu erhalten wird meist eine Kombination aus mehreren der o.g. Methoden benutzt, um die komplette Relevanz für einen Treffer zu bewerten. Die Herausforderung dabei ist die genaue Gewichtung der einzelnen Methoden um die Relevanz eines Treffers optimal zu bewerten. Für jeden Treffer wird dann ein Relevanz-Score berechnet, der sich aus den einzelnen Methoden zusammensetzt. Nach diesem Score wird in einer Liste absteigend sortiert, um das relevanteste Ergebnis als erstes Element zu erhalten.

Sollte sich der Score eines Treffers in der Relation zu anderen Wertungen weit absetzen, kann dieser Treffer dem Nutzer auch direkt vorgeschlagen werden.¹³ Dieses Vorschlagen

¹⁰Vgl. Page, Brin, Motwani und Winograd [10]

¹¹Vgl. Joachims, Radlinski, S. 1 [11]

¹²Vgl. Manning, Raghavan, Schütze [4]

¹³Vgl. Turnbull, Berryman, S. 225-228 [12]

von Ergebnissen kann bereits bei der Eingabe einer Abfrage geschehen, durch sogenannte „Search Completion“.¹⁴

Für kleinere Anwendungen ist hierbei meist ein manuelles Einstellen nach einem Trial-and-Error Verfahren notwendig, bei denen einige wenige Methoden unterschiedlich gewichtet werden. Dies wird dann von Zeit zu Zeit wiederholt, wenn neue Erkenntnisse aus Tests oder dem produktiven Betrieb zurückkommen.¹⁵

Große Suchmaschinen Nutzen hierfür allerdings wie bereits erwähnt hunderte Methoden und evaluieren deren Erfolg im produktiven Betrieb durch proprietäre statistische Methoden.¹⁶

2.4. SOLR

SOLR oder Apache SOLR ist die im vorliegenden Projekt vorliegende Such Engine, die als Web Schnittstelle dient, um auf einer Datenmenge Suchanfragen mit Apache Lucene oder Apache Lucene auszuwerten.¹⁷

Apache Lucene, oder auch kurz Lucene genannt, ist eine mächtige Suchbibliothek, die plattformunabhängig von verschiedenen bekannten Apps, wie z. B. Netflix eingesetzt wird. Lucene nutzt für die Indizierung zu durchsuchender Dokumente Textfelder, wie z. B. „title“ für den Titel eines Dokumentes, um sowohl den Inhalt als Volltext sowie auch die Attribute durchsuchen zu können.¹⁸

SOLR benutzt die hier die Indexing Funktionen von Lucene, um in Echtzeit alle verfügbaren Dokumente zu indizieren um bei einer Suche nur den Index durchsuchen zu müssen. Mit Apache Zookeeper wird dann eine API zur Verfügung gestellt, welche Synchronisierung, Namensregister und die Verteilung der Konfiguration bereitstellt.¹⁹

¹⁴Vgl. Turnbull, Berryman, S. 206-218 [12]

¹⁵Vgl. Zaragoza, Najork, S. 3 [5]

¹⁶Vgl. Taylor, Zaragoza, Craswell, Robertson, Burges [13]

¹⁷Siehe Apache SOLR [14]

¹⁸Siehe Apache Lucene [15]

¹⁹Siehe Apache SOLR [14]

3. Anforderungen und Problemanalyse

Im folgenden Kapitel sollen alle wesentlichen Funktionen der Erweiterung dargestellt werden. Da es sich um ein kleines Projekt mit einem abgestecktem Rahmen handelt und es kein Team gibt, welches die Entwicklung durchführt, wird das Wasserfallmodell für diese Entwicklung verwendet.

3.1. Vorgehensweise

Zur Erfassung der Anforderungen bzw. Requirements Engineering werden User Stories benutzt. Diese sind bekannt aus agilen Softwareentwicklungsmodellen, wie z. B. Scrum, entstanden aber durch praktische Erfahrungen in der Softwareentwicklung. Konzeptioniert wurden sie von Dr. Ivar Jacobsen¹ und Ron Jeffries². Mithilfe einfacher Sprache wird aus der Sicht des Stakeholders das Ziel einer Story in einem kurzen Satz zusammengefasst. Anschließend wird dieses Ziel begründet, um die Wichtigkeit und Existenzberechtigung der Story zu begründen. User Stories sind dabei auch Anforderungen nach dem SMART Prinzip³, da diese nur einen sehr kleinen abgesteckten Teilbereich einer Funktionalität enthalten. Dadurch sind sie einfacher schätzbar, umsetzbar und testbar. Anhand der ermittelten User Stories werden nach der Entwicklung Akzeptanztests durchgeführt, um den Erfolg des Endproduktes objektiv zu bewerten.

3.2. User Stories

Die Anforderungen umfassen alle Aktionen, welche der Nutzer in der Anwendung durchführen will. Ziel aller Anforderungen ist die übergreifende Suche über mehrere Kategorien nach effizient nach mehreren Kriterien zu durchsuchen und zu bewerten. Anhand dieses Ziels werden User Stories entwickelt, in denen ein Nutzer und andere Personen ihre Anforderungen an das zu entwickelnde Produkt stellen. Nichtfunktionale Anforderungen werden bei dieser Anforderungserhebung nicht beachtet, da es sich hierbei um die Erweiterung einer bestehenden API handelt und Punkte wie Benutzerfreundlichkeit und User Experience hierbei wenig relevant sind, beziehungsweise das Entwicklungsumfeld durch die bereits bestehende Anwendung vorgegeben ist. Die Priorität der einzelnen User Stories ergibt sich aus der unten gegebenen Reihenfolge. Ich, als Benutzer, will ...

... für ein gegebenes Suchkriterium relevante Suchergebnisse über mehrere Kategorien hinweg erhalten, damit mit einer einzelnen Suche nur eine geringe Teilmenge der Datenbank angezeigt wird.

¹Vgl. Jacobson, Spence, Kerr 2016 [16]

²Vgl. Ron Jeffries [17]

³Vgl. Witte 2019a, S. 67 [18]

- ... die erhaltenen Suchbegriffe nach Kontext und Wahrscheinlichkeit gewichtet erhalten, damit diese im späteren Verlauf sortiert werden können. Der Kontext ergibt sich aus möglichen Schlagworten, die im Suchbegriff verwendet worden, womit z. B. eine Kategorie, ein Attribut eines Ergebnisses oder höher gewertet wird. Beispiele wären:
 - ... der Titel eines Buches wird „relativ“ genau als Suchbegriff eingegeben, folglich wird dieses Buch stärker gewichtet.
 - ... der Suchbegriff enthält den Term „Video“, folglich werden alle Videos priorisiert.
- ... die erhaltenen Suchbegriffe anhand des errechneten Gewichts absteigend sortiert zurückgegeben werden, damit das relevanteste Suchergebnis auf der Suchergebnisseite ganz oben steht.
- ... ein mit hoher Wahrscheinlichkeit gesuchte Suchergebnis als Vorschlag angezeigt bekommen, damit auf der Suchergebnisseite eine schnelle Navigation zu diesem Ergebnis möglich ist. Dabei soll nur ein Vorschlag angezeigt werden, wenn er der einzige mit einer hohen Wahrscheinlichkeit, in der Suchergebnisliste vorhanden durch das vorher errechnete Gewicht, und dieses Gewicht eine gewisse Schwelle überschreitet. Dadurch sollen verhindert werden, dass von ähnlich relevante Suchergebnisse nur eines vorgeschlagen wird und ein Suchergebnis vorgeschlagen wird, welches für das gegebene Suchkriterium irrelevant ist.

4. Konzeption

4.1. Analyse der Arbeitspakete

4.2. Konzeption

5. Entwicklung des Prototyps

6. Evaluation

7. Diskussion der Ergebnisse

7.1. Abgleich der Ergebnisse mit den gestellten Anforderungen

7.2. Identifizierte Probleme

7.3. Bewertung

8. Ausblick

Anhang

I. Bilder

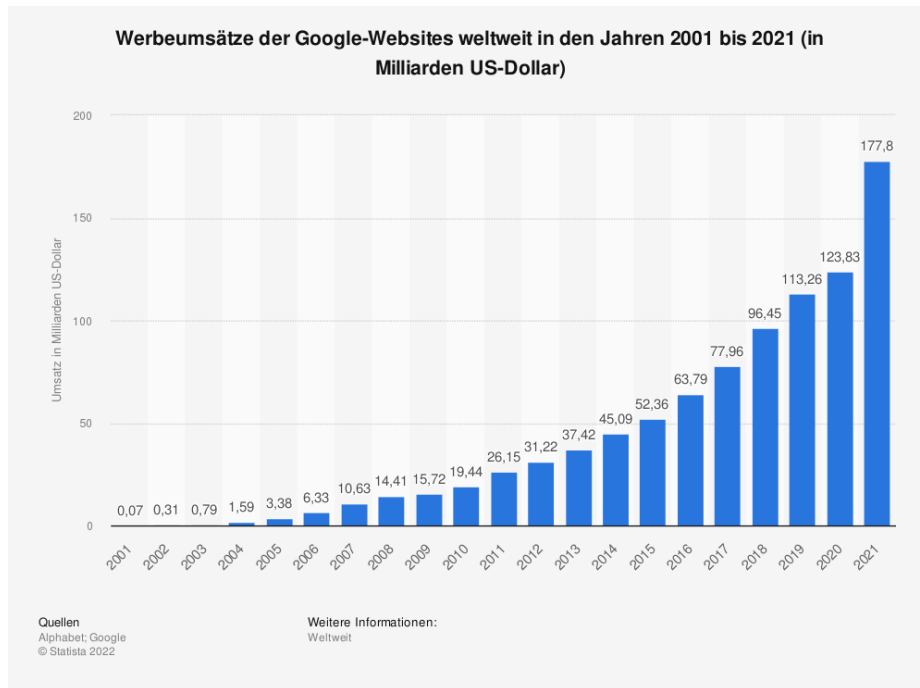


Abbildung 1.: Werbeumsätze Google Websites [19]

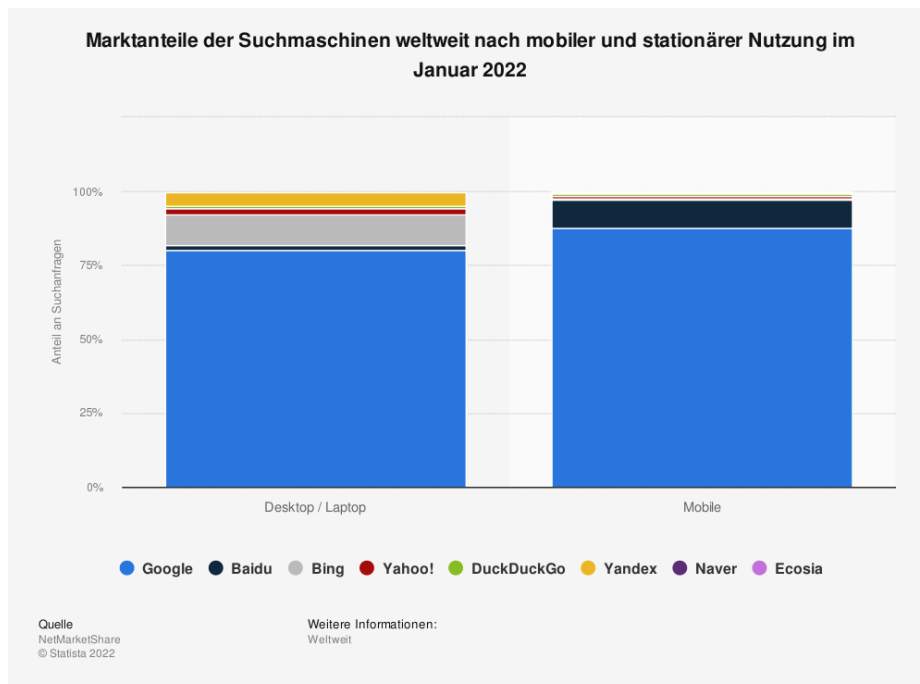


Abbildung 2.: Marktanteil Google [20]

Martin Luther

Bücher, Bibeln & Kommentare

7 Ergebnisse

	MARTIN LUTHER Der kleine Katechismus wie sie ein Hausvater den Seinen einfältig vorhalten soll Ich bin der Herr, dein Gott. Du sollst nicht andere Götter haben...	
	MARTIN LUTHER Von der Freiheit eines Christenmenschen Dem fürsichtigen und weisen Herrn Hieronymo Mühlpfordt, Stadtvogt zu Zwickau, meinem besondern günstigen Freund...	
	MARTIN LUTHER Eine Auslegung des Vater Unsers	
	MARTIN LUTHER Aus der Tiefe rufe ich, HERR, zu dir Das wichtige Thema Gebet ist wie ein goldener Faden in fast alle Werke Martin Luthers eingewoben. In diesem Andachtsbuch wir...	
	MARTIN LUTHER Der große Katechismus Nach der Fassung des deutschen Konkordienbuches (Dresden 1580)	

7 Suchergebnisse in Bücher, Bibeln & Kommentare

ALLE BÜCHER (7)

Sonstige

21 Ergebnisse

	MARTIN LUTHER Von der Gerechtigkeit	
--	---	--

Abbildung 3.: *Crossload* [2]

Abbildungsverzeichnis

1. Werbeumsätze Google Websites [19]	A
2. Marktanteil Google [20]	A
3. Crossload [2]	B

Tabellenverzeichnis

Literaturverzeichnis

- [1] Duden, “Duden | Wie schreibt man „googeln“? | Rechtschreibung.” <https://www.duden.de/rechtschreibung/googeln>, 2022.
- [2] D. Pfeiderer, “CROSSLOAD.” <https://crossload.org/info/ueber>, Sept. 2022.
- [3] A. Bookstein, “Relevance,” *Journal of the American Society for Information Science*, vol. 30, pp. 269–273, Sept. 2007.
- [4] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. New York: Cambridge University Press, 2008.
- [5] H. Zaragoza and M. Najork, “Web Search Relevance Ranking,” in *Encyclopedia of Database Systems* (L. Liu and M. T. Özsu, eds.), pp. 4650–4655, New York, NY: Springer New York, 2018.
- [6] D. Liu, X.-s. Hua, M. Wang, H.-j. Zhang, and L. Yang, “WWW 2009 MADRID! Track: Rich Media / Session: Tagging and Clustering Tag Ranking *,” 2009.
- [7] G. Developers, “Image labeling | ML Kit.” <https://developers.google.com/ml-kit/vision/image-labeling>, 2022.
- [8] G. Photos, “Google Photos.” <https://www.google.com/photos/about/>, 2022.
- [9] M. Marchiori, “The quest for correct information on the Web: Hyper search engines,” *Computer Networks and ISDN Systems*, vol. 29, pp. 1225–1235, Sept. 1997.
- [10] L. Page, S. Brin, R. Motwani, and T. Winograd, “The PageRank citation ranking: Bringing order to the web.,” Technical Report 1999-66, Stanford InfoLab / Stanford InfoLab, Nov. 1999.
- [11] T. Joachims and F. Radlinski, “Search Engines that Learn from Implicit Feedback,” *Computer*, vol. 40, pp. 34–40, Aug. 2007.
- [12] D. Turnbull and J. Berryman, *Relevant Search: With Applications for Solr and Elasticsearch*. Manning, 2016.
- [13] M. Taylor, H. Zaragoza, N. Craswell, S. Robertson, and C. Burges, “Optimisation methods for ranking functions with multiple parameters,” in *Proceedings of the 15th ACM International Conference on Information and Knowledge Management - CIKM '06*, (Arlington, Virginia, USA), p. 585, ACM Press, 2006.
- [14] Solr, “Apache Solr Reference Guide :: Apache Solr Reference Guide.” <https://solr.apache.org/guide/solr/latest/>, 2022.

- [15] A. Lucene, “Welcome to Apache Lucene.” <https://lucene.apache.org/index.html>, 2022.
- [16] I. Jacobson, I. Spence, and B. Kerr, “Use-Case 2.0: The Hub of Software Development,” *Queue*, vol. 14, pp. 94–123, Jan. 2016.
- [17] R. Jeffries, “Essential XP: Card, Conversation, Confirmation.” <https://ronjeffries.com/xprog/articles/expcardconversationconfirmation/>, 2022.
- [18] F. Witte, *Testmanagement und Softwaretest*. Wiesbaden: Springer Fachmedien, 2016.
- [19] Alphabet, “Werbeumsätze der Google-Websites weltweit in den Jahren 2001 bis 2021.” <https://de.statista.com/statistik/daten/studie/75181/umfrage/werbeumsatz-der-google-websites-seit-2001/>, Feb. 2022.
- [20] NetMarketShare, “Marktanteile der Suchmaschinen - Mobil und stationär 2022.” <https://de.statista.com/statistik/daten/studie/222849/umfrage/marktanteile-der-suchmaschinen-weltweit/>, Feb. 2022.