



## ANFORDERUNGSANALYSE UND MACHBARKEITSTUDIE EINER UNTERNEHMENS-SPEZIFISCHEN SUCHFUNKTION

**LUKAS KIELMANN (BACHELORSTUDIUM INFORMATIK)**

Betreuer: Prof. Dr. Kai Höfig, Prof. Dr. Gerd Beneken

### Motivation

E-Commerce und der D2C-Absatz für produzierende Unternehmen werden in einer immer stärker vernetzten Welt rasant wichtiger. Ein zentraler Bestandteil der Customer-Journey durch einen Webshop ist die Suchfunktion. Laut einer Studie der User Experience Plattform Nosto und der Marktforschungsplattform Censuswide nutzten 69% der befragten Verbraucher beim Online-Shopping als erstes die Suchfunktion. [1] Ziel der Bachelorarbeit ist es die Anforderungen an eine Suchfunktion aus Sicht von Endkunden und firmeninternen Experten zu erfassen und basierend auf diesen einen Prototypen zu erstellen, der mit dem Altsystem verglichen wird. Das Altsystem wurde 2016 mit Elasticsearch 2 aufgebaut und seitdem nicht mehr angepasst. Die zu Verfügung stehenden Inhalte und die Möglichkeiten moderner Volltext-Suchmaschinen haben sich verändert und können bisher nicht berücksichtigt werden. [2]

### Stand der Technik

Auf Volltext-Suchen spezialisierte Suchmaschinen-Frameworks wie Elasticsearch speichern die Dokumente in invertierten Indexen als JSON-Objekte. In diesen wird für jedes indizierte Wort die Dokumente, in denen es vorkommt, sowie die Position im Dokument selbst gespeichert. Zusätzlich kann noch die Häufigkeit und die Position des Wortes im jeweiligen Dokument als Metrik erfasst werden. [2] Mit dem BM25-Algorithmus kann die Relevanz eines Dokumentes für einen Suchbegriff bestimmt werden. Der Algorithmus besteht aus den Teilen Term Frequency TF (Je häufiger ein Begriff in einem Dokument vorkommt, desto wichtiger ist er) und Inverse-Document-Frequency IDF (Je häufiger ein Begriff über alle Dokumente vorkommt, desto unwichtiger ist er)

sowie einiger Konstanten. Je niedriger die Relevanz, desto weniger stimmt das Ergebnis mit dem Suchbegriff überein. Eine hohe Relevanz bedeutet aber nicht, dass ein Ergebnis richtig im Zusammenhang mit der Suchanfrage ist. [3]

### Anforderungsanalyse

Die Anforderungsanalyse wird in vier Schritten durchgeführt. Die multidimensionale Analyse ermöglicht es die Anforderungen der Experten und die Erfahrungen der Kunden zusammenzubringen. Außerdem ist keine Dokumentation über die Einführung des Altsystems mehr vorhanden, mit der frühere Entscheidungen erklärt werden können.

Zunächst wird in einer qualifizierten Expertenbefragung die Sichtweise von elf Mitarbeitern aus den Marketing- und Produktmanagement-Abteilungen der deutschen, französischen und australischen Organisation mit einem Leitfadenterview erfasst. Die Ergebnisse zeigen, dass den Experten es am wichtigsten ist, dass die Suchfunktion Suchvorschläge ausgibt. Diese werden auf der Website unterhalb des Suchfeldes gelistet.

Die befragten Experten gehen davon aus, dass die Kunden ihre Suche abbrechen, falls es für einen Begriff kei-

nen Vorschlag gibt. Etwas weniger wichtig sind alternative Suchergebnisse für Produktabmessungen (z.B. statt "63 x 56 mm" "63,2 x 55,5" mm Etiketten) und eine Auto-Vervollständigung in der Suchfunktion (engl. Typeahead). Darüber hinaus geben die Experten Listen mit Suchbegriffen ab, die für sie relevant sind. Mit diesen werden im späteren Verlauf das Altsystem und der Prototyp verglichen. Die Artikelnummer, der Produktname und vordefinierte Produktanwendungen sind für die Experten die wichtigsten Elemente des Index. Alle weiteren Daten sind nicht unbedingt nötig und sollten eine geringere Prioritäten haben.

Die zweite Dimension umfasst eine quantitative Endkundenbefragung mit einem Online-Fragebogen. Die Befragung wird über einen Newsletter unter 8607 aktiven Kunden von Avery Zweckform verteilt. 150 Personen nehmen an der Befragung teil von denen 98 diesen abschließen. Sie dient dazu zu messen, wonach und mit welchen Begriffen Endkunden suchen. Mit den gesammelten Suchbegriffen soll die Liste der Suchbegriffe der Experten erweitert werden. Die Ergebnisse zeigen, dass fast 51% der Teilnehmer die Suchfunktion im letzten Monat verwendet haben und mit ihrer Funktion zufried-

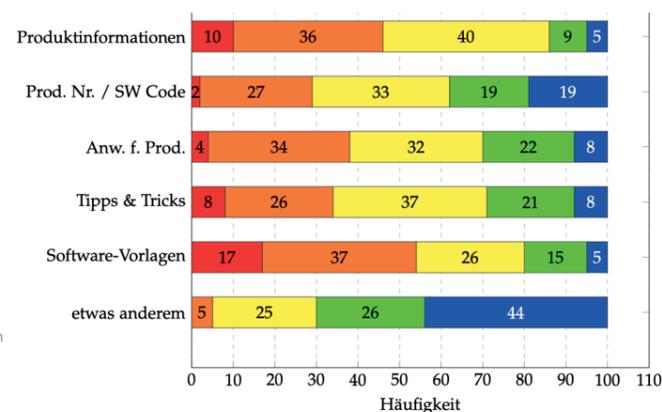


Abb. 1: Endkunden suchen nach... (100 Teilnehmer)

# ROSENHEIMER INFORMATIKPREIS INF-BACHELOR

den sind. Wie in Abb. 1 zu sehen ist, suchen die Endkunden am häufigsten nach Software-Vorlagen für Blanko Etiketten, Produktinformationen und Tipps & Tricks Seiten, die den Nutzern Inspirationen geben sollen. Sie suchen am häufigsten mit Produktnummern oder Software-Codes, Produktanwendungen und Produktabmessungen. Als Produktanwendungen werden alle Begriffe bezeichnet, die mit einer Produkteigenschaft kombiniert sind, z.B. Temperaturbeständige Etiketten, oder von Produktmanagern vordefinierte Anwendungen wie z.B. Honigetiketten. Als dritter Aspekt wird das Altsystem untersucht und so Schwachstellen für eine Drop-In Ersatzlösung identifiziert. Da Open Search seine Code Basis mit Elasticsearch 7.10 teilt, sind die Systeme kompatibel miteinander. Durch den Versionssprung von Elasticsearch selbst ist das Mapping und die Suchvorlage dennoch inkompatibel. Für die Machbarkeitsstudie müssen beide Teile aktualisiert werden. Auch fällt auf, dass lediglich acht der 21 von der Firma gepflegten Seitenarten indiziert werden. Eine Produktdetailseite enthält andere Attribute als eine Seite für Produktlisten oder eine Seite für Ideen & Inspirationen. Seiten, die nicht indiziert sind, können in der Suchfunktion nicht gefunden werden. Die vierte Dimension bildet die Analyse der Google Analytics Daten der deutschen und englischen Websites. Ziel ist es Gruppierungen im Suchverhalten zu finden und Top-Suchbegriffe zu identifizieren. Es werden jeweils die Top 1000 Begriffe (nach der Häufigkeit absteigend sortiert) analysiert. Zu sehen ist, dass auf der deutschen Website zu fast 90,68 % mit Artikelnummern gesucht wird. Auf der englischen Website ist das Verhalten anders. Dort wird nur mit 51,25 % mit Artikelnummern gesucht. Darauf folgen Produktformen mit 23,91 % und Produktanwendungen mit 10,49 %. Daraus lässt sich ableiten, dass die Märkte unterschiedliche Anforderungen an die Suchergebnisse haben. Für den Test des Prototyps werden jeweils die Top 10 Artikelnummern und nicht Artikelnummern ausgewählt.

Es lassen sich vier Personas ableiten, die besonders von dem Projekt betroffen sind. Beneken et al. schreiben, dass Personas eine Modellierung der Nutzergruppen eines Produkts sind. Sie dienen dazu, die Nutzer besser zu verstehen. Das Projektteam kann sich mithilfe der Modelle eigene Erwartungen mit der Nutzerperspektive abgleichen. [4] Aus der Gruppe der Experten lassen sich die Personas Marketing Manager, Vertriebsmitarbeiter und Consumer Center Agent ableiten. Der Endkunde ist die vierte Persona. Aus den gesammelten Informationen werden mit den Personas 24 Anforderungen nach dem User Story Schema formuliert. Eine User Story verbindet eine Persona mit einer Anforderung und einem Zweck. Neben dem Zweck werden über Akzeptanzkriterien zusätzlicher Kontext für die Entwickler hinzugefügt.

### Implementierung und Bewertung des Prototyps

Die Machbarkeitsstudie prüft welche Anforderungen mit einer Drop-In Ersatzlösung umgesetzt werden können und wie sich die Qualität der Ergebnisse verändert. Zunächst wird die bestehende Suchvorlage und das Index-Schema aktualisiert, damit es mit OpenSearch 2.10 kompatibel ist. Daraufhin werden die Anforderungen nacheinander geprüft und die dafür benötigten Änderungen am Index und der Suchvorlage implementiert. Um durch Änderungen eine Verschlechterung der Ergebnisse zu vermeiden, wird nach jedem Schritt die Relevanz der Ergebnisse für einen Suchbegriff als Wert mit dem BM25-Algorithmus berechnet. Zur Vergleichbarkeit der Ergebnisse werden die Suchergebnisse der 74 gesammelten Testsuchbegriffe für das Altsystem und den Prototypen erfasst. Die indizierten Seitentypen werden in die fünf Kategorien ideas\_inspiration, custom\_printing, product, template und help gruppiert. Aus jeder Kategorie werden initial vier Ergebnisse auf der Website präsentiert. In der Bewertung wird jeweils das Top Ergebnis jedes Begriffs und jeder Kategorie untersucht

(370 Ergebnisse). Ein Ergebnis kann besser, gleichwertig oder schlechter als das Altsystem sein. Ein gleichwertiges Ergebnis ist entweder gleich gut oder gleich schlecht. Ein schlechteres Ergebnis ist entweder kein Ergebnis oder ein falsches Ergebnis. Die Analyse zeigt, dass der Prototyp 100,75% mehr Ergebnisse liefert als das Altsystem. In 26,75 % handelt es sich um ein besseres Ergebnis als das Altsystem. In 100 von 370 Fällen haben beide Systeme keine Suchergebnisse. Dies zeigt eine Datenlücke und ist besonders in der Gruppierung help ausgeprägt.

### Fazit und Ausblick

Die Machbarkeitsstudie zeigt, dass eine Aktualisierung der Suchmaschine bereits deutlich bessere Ergebnisse liefert. Die neue Lösung wird für die 15 Webshops von Avery Zweckform ausgerollt. Die unerfüllten Anforderungen könnten durch eine Suchfunktion basierend auf künstlicher Intelligenz erfüllt werden. Als kurz bis mittelfristige Maßnahmen zur Verbesserung der Suchfunktion sollten Systeme zur Überwachung und Auswertung des Nutzerverhaltens und der Suchergebnisse implementiert werden. Mit den gewonnenen Daten können automatisierte Tests eingerichtet werden, die bei Datenlücken auf der Website die Marketing-Manager informieren.

### Bibliography

- [1] Nostos, Censuswide, The Future of Ecommerce Search, 2023, p. 2.
- [2] A. Mahapatra and S. Biswas, "Inverted indexes: Types and techniques," International Journal of Computer Science Issues, vol. 8, no. 4, pp. 384-392, 07 2011.
- [3] S. Connelly, "Practical BM25 — Part 2: The BM25 Algorithm and its variables," Elasticsearch, 19 April 2018. [Online]. Available: <https://www.elastic.co/blog/practical-bm25-part-2-the-bm25-algorithm-and-its-variables>. [Accessed 17 12 2024].
- [4] G. Beneken, F. Hummel and M. Kucich, Grundkurs agiles Software-Engineering: Ein Handbuch für Studium und Praxis, Springer Fachmedien Wiesbaden und divibib GmbH, 2023, pp. 129-131.