

# Suchen und Ranken von klinischen Leitlinien unter besonderer Berücksichtigung von Synonymen für die Qualität des Suchergebnisses

Adrian Dröse

2021

## 1 Einleitung

Das Problem dieser Bachelorarbeit umfasst den Themenbereich der klinischen Leitlinien in der Medizin. Diese Leitlinien geben dem Leser eine Auskunft über den aktuellen Erkenntnisstand in Bezug auf verschiedene Behandlungsmöglichkeiten bei bestimmten Krankheitsbildern. Sie sollen bei einer Entscheidungsfindung in der klinischen Versorgung unterstützen[1]. Die AWMF (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V.) veröffentlicht auf ihrer Webseite die Leitlinien deutscher medizinischer Fachgesellschaften in einem qualitätsgesicherten Prozess. Dabei wird eine Anzahl von ungefähr 800 Leitlinien bereitgestellt, welche mit einer Suchfunktion zu finden sind. Das genaue Problem der Bachelorarbeit befasst sich mit dieser Suchfunktion, denn die gelieferten Ergebnisse sind oftmals nicht den Ansprüchen der AWMF entsprechend. Aktuell sieht eine typische Anfrage des Nutzers wie folgt aus: Der Benutzer verwendet die Suchfunktion, um eine der Anfrage entsprechende Leitlinie zu erhalten. Zum eingegebenen Suchbegriff werden über die Volltextsuche eines relationalen Datenbankmanagementsystems, in dem die Leitlinientexte hinterlegt sind (nach Extraktion aus den PDF-Dokumenten, in denen die Leitlinien derzeit üblicherweise veröffentlicht werden), mittels exaktem Stringmatching eben jene Leitlinien gefunden, die genau den Suchbegriff enthalten. Die Kriterien für das Ranking der so erzeugten Suchergebnisse sind derzeit nicht bekannt. Diese so erzeugte Ausgabe ist nach den Angaben der AWMF nicht gut, da oft verschiedene Leitlinien gefunden und an prominenter Stelle gelistet werden, die für

die Suchanfrage thematisch nicht relevant sind. Gleichzeitig werden oft relevante Ergebnisse gar nicht gefunden, da sie nicht den Suchbegriff selbst einhalten sondern eines seiner (in der Medizin oft vielfältigen) Synonyme. Diese Bachelorarbeit soll erforschen, ob und in welchem Umfang die derzeitige Suchfunktion durch den Einsatz der folgenden Methoden für den genannten Einsatzbereich verbessert werden kann:

- Unschärfe Suche und verbessertes Ranking durch die Nutzung einer auf die Volltextsuche spezialisierten Indexierungs- und Suchmaschine (Elastic Search oder Solr)
- Nutzung von Synonymlisten zur Erweiterung des Suchraumes für die jeweils gesuchte Entität. Die Synonyme sollen dabei aus gängigen medizinischen Klassifikationen gewonnen werden, die zu verschiedenen Entitäten jeweils mehrere Synonyme und Bezeichnungen enthalten (ICD-10, OPS, LOINC) [2]

Die Evaluation der Ergebnisse soll auf zweierlei Weise erfolgen: Zunächst wird mithilfe vorhandener Listen themenverwandter Leitlinien (s.u. Abschnitt Evaluation) ein systematischer Vergleich des Einflusses von Synonymlisten auf das Suchergebnis durchgeführt. Dabei soll auch untersucht werden, ob der Einfluss verschiedener Synonymlisten auf die Ergebnisqualität von der medizinischen Fachrichtung der jeweiligen Leitlinien abhängt. Anschließend soll zu einer Auswahl von 10 Suchbegriffen, die von der AWMF vorgegeben werden in direkter, darstellungsneutraler Gegenüberstellung durch die AWMF bewertet werden, ob jeweils das alte, mit der bisherigen Suchfunktion erstellte Suchergebnis, oder jenes Suchergebnis besser ist, welches mit eben der neuen Methode erstellt wurde, die sich in der ersten systematischen Analyse als die Beste herausgestellt hat.

## 2 Methoden und Vorgehensweise

Um die erforderlichen Daten (Leitlinien und zugehörige Listen themenverwandter Leitlinien) von der AWMF-Webseite herunterzuladen, wird ein Web Scraper geschrieben [3]. Hierbei wird jeweils die Langfassung der gegebenen Leitlinien verwendet, da diese Version in jedem Fall gegeben ist. Für jede Leitlinie wird auch die Referenzliste themenverwandter Leitlinien für die spätere Evaluation

extrahiert (vgl. Abschnitt 3.1). In einem nächsten Schritt wird der Leitlinientext aus der PDF-Datei extrahiert und anschließend indexiert.[4].

Die Indexierung erfolgt mithilfe der Suchmaschinen Solr oder Elasticsearch. Diese Suchmaschinen vereinfachen das Arbeiten mit großen Dokumentenmengen und das Finden von Dokumenten zu Suchbegriffen. Beide Engines basieren auf Apache Lucene, einer Programmibibliothek für Volltextsuche [5].

Die gewählten Engines bieten die Möglichkeit für die Berücksichtigung von Synonymlisten des Suchbegriffes. Während die Dokumente von einer der genutzten Engines (Suchmaschinen) indexiert werden, also der Suchindex gebaut wird, werden die Texte bereits vorverarbeitet. Bei dieser Vorverarbeitung werden Methoden aus dem Bereich des Information Retrieval verwendet, wie beispielsweise stop word removal und stemming [6, 7]. Diese Methoden werden durch die Engine erledigt, weshalb eine Implementierung solcher Funktionen nicht notwendig ist.

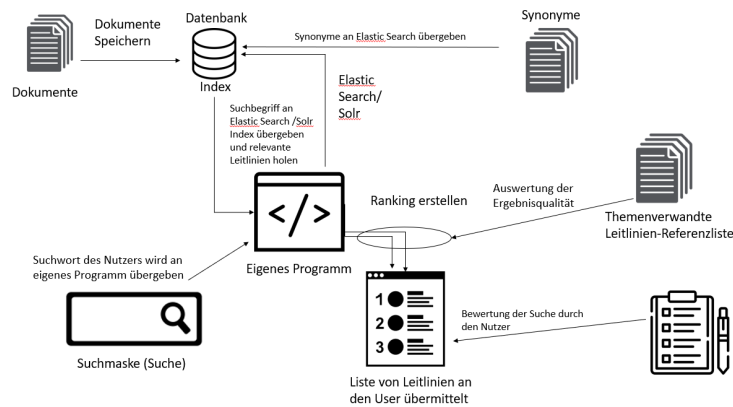


Abbildung 1: Bau des Systems

In Figure 1 wird visuell der Aufbau des Systems beschrieben und soll die grundlegende Vorgehensweise darstellen.

Damit die jeweiligen Synonyme verwendet werden können, werden Synonymlisten aus 3 im deutschsprachigen Raum verbreiteten medizinischen Klassifikationen erstellt. Diese sind ICD-10(International Statistical Classification of Diseases and Related Health Problems), OPS (Operationen- und Prozedurenschlüssel) und LOINC (Logical Observation Identifiers Names and Codes) ver-

wendet. Für LOINC existiert RELMA (Regenstrief LOINC Mapping Assistant), wodurch die in LOINC primär englischsprachigen Bezeichnungen in weiteren Sprachen wie Deutsch zur Verfügung stehen. Durch diese Synonyme ist ein höherer Recall von relevanten Ergebnissen zu erwarten. Diese Synonyme und der geforderte Suchbegriff werden anschließend genutzt, um mittels einer der gegebenen Suchmaschinen die Dokumentendatenbank zu durchsuchen. Als Ergebnis wird eine Liste von entsprechenden Leitlinien geliefert, deren Qualität in dieser Arbeit evaluiert werden soll. Das Eigene Programm, welches in dieser Arbeit verwendet wird, wird ein Auswertungstool sein, das zu den Leitlinien Suchbegriffe definiert, mit diesen automatisiert nach anderen Leitlinien sucht und anschließend eine Auswertung mithilfe der Listen themenverwandter Leitlinien durchführt (siehe Abschnitt 3, Evaluation).

Damit die AMWF-Mitarbeiter sich im zweiten Teil der Evaluation für ein geeignetes Ergebnis entscheiden können, wird eine Gegenüberstellung durchgeführt, bei dem zwei Ergebnislisten aufgeführt werden. Die eine Liste wird dabei die ursprüngliche Ausgabe sein, also die Liste der bisherigen Leitliniensuchfunktion der AWMF und die andere Liste wird die neue Ausgabe sein, die im Laufe dieser Arbeit erforscht und entwickelt wird. Das ursprüngliche System nutzt dabei die Volltextsuche von MySQL und verwendet keine Betrachtung von Synonymen.

## 3 Evaluation

### 3.1 Bewertung der Suche nach themenverwandten Leitlinien

Der Nutzer soll die Suchfunktion nutzen können und erhält bei der Eingabe seines Suchbegriffes eine Auflistung der relevantesten Ergebnisse, wobei die Relevanz mit der Hilfe der AWMF-Mitarbeiter und der Gegenüberstellung im zweiten Teil der Evaluation bestimmt wird. Sollte der Nutzer nun eine der vorgeschlagenen Leitlinien auswählen, werden ihm anschließend zu der ausgewählten Leitlinie themenverwandte Leitlinie aufgelistet. Um zu jeder Leitlinie auch themenverwandte Leitlinien zu finden, werden die Titel genutzt. Hierbei werden die ausschlaggebenden Suchbegriffe aus dem Titel von Leitlinie A herausgesucht und anschließend verwendet, um Synonyme oder ähnliche Begriffe in allen Leitlinien B zu finden. Für die genaue Definition dieser ausschlaggebenden Suchbegriffe existieren mehrere Möglichkeiten. Die erste Variante würde den gesamten Titel

verwenden und mit der Hilfe einer der eingerichteten Suchmaschinen die themenverwandten Leitlinien heraussuchen lassen. Für die zweite Variante werden die Synonymlisten genutzt und dem Titel werden medizinische Fachbegriffe entnommen, welche in den Synonymlisten enthalten sind.

Nun ist die Frage: Wie wird das Ergebnis in Bezug auf die Listen der themenverwandten Leitlinien ausgewertet? Für die Ergebnisliste themenverwandter Leitlinien wird bestimmt, wie viele der ersten  $K$  von der Suchmaschine zurückgegebenen Trefferleitlinien mit den Leitlinien in der jeweiligen Referenzliste themenverwandter Leitlinien übereinstimmen. Genauer gesagt wird dabei der Recall-at- $k$ , also der Recall innerhalb der Top- $K$ -Ergebnisse bestimmt. Zusätzlich wird für die Erstellung der Listen der unterschiedliche Gebrauch der 3 medizinischen Klassifikationen für die Synonymexpansion der Suche in der Suchmaschine genutzt. Dadurch könnte jeweils eine Liste ausgegeben werden, die nur eine der 3 medizinischen Klassifikationen nutzt und auch eine weitere Liste, die alle 3 Klassifikationen verwendet. So soll bezüglich dieser Klassifikationen ermittelt werden:

- welche der Klassifikationen am besten für diesen Zweck funktioniert
- ob Kombinationen mehrerer Klassifikationen besser funktionieren als einzelne Klassifikationen (Wenn ja, welche wären es?)
- ob es dabei erkennbare Unterschiede zwischen den Fachrichtungen gibt

Die Ergebnislisten hätten dann eine Länge bis zu 10 Leitlinien (also  $k=10$ ). Das Rankingsystem der jeweiligen Suchmaschine hilft hierbei unterstützend, um ein Ranking dieser Top 10 zu erstellen. Um diese Bestenliste zu ermitteln, wird eventuell eine Konfiguration des Rankings notwendig sein. Dieser Schritt ist abhängig davon, wie gut die Suchergebnisse von vornherein sind.

### **3.2 Bewertung der Suche durch den Nutzer**

In diesem Abschnitt wird untersucht, ob die Stichwortsuche gut funktioniert und ob sie den Anforderungen der AWMF entsprechende Ergebnisse ausgibt. Für die Bewertung der Suche durch den Nutzer wird die zuvor erwähnte Gegenüberstellung der Ergebnisse verwendet. Die Auswertung erfolgt anhand von 10 Suchbegriffen, die von der AWMF vorgegeben werden. Für jeden dieser 10 Suchbegriffe wird sowohl eine Top-10 Liste mit Ergebnissen des bisher bei der

AWMF betriebenen Systems als auch eine Top-10 Liste mit Ergebnissen des neuen Systems ausgegeben. Die AWMF-Mitarbeiter werden dann für jeden Suchbegriff abstimmen, welche der jeweiligen Top-10 Ergebnislisten besser ist. Anhand dieser Beurteilung soll ermittelt werden, ob das Ziel der Verbesserung der Ergebnisqualität durch die vorgenommenen Maßnahmen erreicht werden konnte. Während der Abstimmung soll es für die Bewertenden nicht ersichtlich sein, welchem System die jeweilige Liste entstammt.

## Literatur

- [1] M Lelgemann u. a. “Leitlinien”. In: *Bundesgesundheitsblatt Gesundheitsforsch. Gesundheitsschutz* 48.2 (2005), S. 215–220.
- [2] “ICD-10-GM Version 2020, Systematisches Verzeichnis, Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision, Stand: 20. September 2019”. In: (2019).
- [3] Ryan Mitchell. *Web scraping with Python: Collecting more data from the modern web*. Ö'Reilly Media, Inc., 2018.
- [4] Akshay Kulkarni u. a. “Extracting the Data”. In: *Natural Language Processing Recipes*. Springer, 2019, S. 1–35.
- [5] Ugur Kılıç u. a. “Comparison of Solr and Elasticsearch Among Popular Full Text Search Engines and Their Security Analysis”. In: *Future Internet of Things and Cloud Workshops, 2015 6th International Conference on*. 2016, S. 163–168.
- [6] Jaideepsinh K Raulji u. a. “Stop-word removal algorithm and its implementation for Sanskrit language”. In: *International Journal of Computer Applications* 150.2 (2016), S. 15–17.
- [7] Cristian Moral u. a. “A survey of stemming algorithms in information retrieval.” In: *Information Research: An International Electronic Journal* 19.1 (2014), n1.