

## Handreichungen

### Wie suche ich mit der Suchmaschine des Hochschulbibliothekszentrums?

Die Suchfunktionalität ist grundsätzlich an bekannte Suchmaschinen wie Google angelehnt und sollte ohne Einweisung zu benutzen sein. In dieser Handreichung werden wesentliche Eigenschaften dokumentiert und einige Besonderheiten beschrieben.

#### Einfache Wörter aus dem Titel

Auf die Groß- und Kleinschreibung kommt es nicht an. Interpunktion wird ignoriert.

- Erziehung
- Technik
- Wirtschaft

Werden Wörter aus dem Titel gesucht, so werden auch in vielen Fällen Wortformen einbezogen, die eine verkürzte Form des Suchwortes sind.

- technik → techn

Die Verkürzung (stemming) ist sprachspezifisch auf die deutsche Sprache eingestellt und löst deswegen die meisten Singular- und Pluralformen auf. Dabei wird ein Algorithmus eingesetzt, der Wörter auch unerwünscht stark verkürzen kann. Daher werden immer auch alle unverkürzten Wortformen mitgespeichert. Das Stemming kann dazu führen, dass die Suche nach einer verkürzten Form weniger Treffer als die Suche nach dem vollständigen Begriff ergibt.

Umlaute können in Original (ä,ö,ü,ß) oder aufgelöst (ae,oe,ue,ss) oder in Grundform (a,o,u) eingegeben werden.

#### Eigennamen

Eigennamen (Personen) werden nicht auf ihre Stammform zurückgeführt. Sie werden nur gefunden, wenn das Suchwort mit dem Wort im Index übereinstimmt. Die Reihenfolge von Vor- und Nachname in der Suchanfrage spielt keine Rolle.

- Johann Wolfgang von Goethe
- Goethe, Johann Wolfgang

Um das Stemming des Suchbegriffs zu verhindern, muss im Autorenfeld recherchiert werden. Ansonsten kann nicht erkannt werden, dass es sich um einen Eigennamen handelt!

Eigennamen sind mit der GND verknüpft. Die Suche nach "Tiger" führt also auch zu Treffern mit Tucholsky

## Verknüpfung mehrerer Wörter

Je mehr Suchwörter eingegeben werden, desto genauer und schneller ist die Suche und desto kleiner die Treffermenge, aber auch desto größer die Wahrscheinlichkeit, keine Treffer zu erzielen. Alle eingegebenen Suchwörter werden im Normalfall mit der booleschen AND-Operation verknüpft, wodurch die Treffermenge eingeschränkt wird. Es wird nur noch die Schnittmenge aller Treffermengen der Einzelwörter ausgegeben.

- lexikon harenberg 1994 → 1 Treffer
- lexikon literatur → 1819 Treffer
- lexikon weltliteratur → 258 Treffer
- lexikon literatur weltliteratur → 97 Treffer

Werden zwei Wörter mit einem vertikalen Bindestrich (|) verknüpft, so findet eine Ausweitung der Suche nach dem einem oder dem anderen Wort statt (OR Operation).

- lexikon | weltliteratur → 52357 Treffer

Wird ein Wort mit einem vorangestellten Negationszeichen (-- ) versehen, so werden alle Treffer, die dieses Wort enthalten, aus der Treffermenge entfernt (NOT-Operation).

- lexikon literatur -weltliteratur → 1722 Treffer

Mit runden Klammern wird die Präzedenz der Operatoren bestimmt.

- (lexikon | weltliteratur) literatur → 3385 Treffer

## Phrase

Eine Phrase (Wortfolge) wird in doppelte Anführungszeichen gesetzt. Es werden nur Treffer gefunden, die genau diese Wortfolge enthalten. Mit einer direkt nachgestellten Tilde kann die Anzahl der Wörter angegeben werden, die zwischen den Einzelwörtern der Phrase vorhanden sein dürfen.

- "lexikon der weltliteratur" → 8 Treffer
- "lexikon weltliteratur"~1 → 170 Treffer

## Komposita

Komposita (Bindestrichwörter) werden auch in der zusammengeführten Form indexiert und gesucht, um die Treffermenge auszuweiten. Beispiel: eine Suche nach dem Wort E-Book erweitert sich automatisch um die

Suche nach "E oder Book". Eine Suche nach Technologietrends findet Dokumente mit Technologietrends und Technologie-Trends

## Trunkierung

Mit der Rechtstrunkierung können Wörter gefunden werden, deren linker Wortteil mit dem Suchwort übereinstimmt.

Beispiel: wirtschaft\* findet Treffer mit wirtschaft, wirtschaften, wirtschaftswissenschaft usw.

Hinweis: die Wortteile werden ebenfalls auf einen Wortstamm verkürzt und erst danach eine Suche über alle gebildeten Wortteile durchgeführt, die im Index vorhanden sind. Da sich im Index die Wörter um Kompositaformen erweitern, kann es passieren, dass Trunkierungen auch auf diese Wörter angewendet werden.

## Vergrößerung der Treffermenge

Neben der OR-Operation gibt es weitere Möglichkeiten, Treffermengen zu vergrößern. Die Suchmaschine besitzt keinen Stoppwort oder Antiphrasenfilter (Filter zum Aussortieren unerwünschter Wörter). Wenn bei der Suche auf Stoppwörter und Füllwörter sowie Wörter mit unsicherer Schreibweise verzichtet wird, so kann die Treffermenge vergrößert werden.

- lexikon der weltliteratur → 215 Treffer
- lexikon weltliteratur → 258 Treffer

Die Suchmaschine verwendet automatisch bibliothekarische Normdateien wie die Gemeinsame Normdatei (GND), um verschiedene Schreibweisen von Personen, Körperschaften, Geografika und Schlagwörtern abzudecken und auf diese Weise Treffermengen auszuweiten.

- hildegard von bingen
- hildegardis bingensis

## Verkleinerung der Treffermenge

Neben der AND-Operation gibt es weitere Möglichkeiten, Treffermengen zu verkleinern. Die Suchmaschine besitzt dafür spezielle Mechanismen für das automatische Erkennen von Standardnummern und für das Bilden von Facetten. Bei der Standardnummernsuche ist das erwartete Ergebnis sehr klein (meistens nur ein Treffer) und bei einer Facettenbildung werden gemeinsame Eigenschaften von Dokumenten erkannt und in Kurzform für eine vertiefende Suche angeboten. Beispiel ISBN-Suche

- 978-1-4724-1152-5 [javascript:][javascript:]

- 9781472411525 [javascript:][javascript:]

Bei der ISBN-Suche wird eine gültige ISBN-13 oder ISBN-10 erkannt, dabei kommt es auf die Bindestriche nicht an. Diese können auch weggelassen werden. Gefundene Treffermengen können je nach gefundenen Attributen der Treffer in Teilmengen (Facetten) aufgeteilt werden. Damit ist eine schnelle Übersicht über die gefundenen Treffer möglich, auch über diejenigen, die nicht auf der ersten Seite zur Anzeige gelangen.

## **Ranking und Sortierung**

Das Relevanzranking ist in der einfachen Suche, die über alle Felder durchgeführt wird, aktiv. Die bibliografischen Daten sind unter Benutzung des probabilistischen Ranking-Algorithmus Okapi BM25 ("best match number 25") indexiert, der aus den 80er Jahren stammt und von der London City University in Zusammenarbeit mit der British Library entwickelt und auf Grundlage von MARC-Daten getestet wurde. Trifft die Suche auf passende Wörter, wird dies belohnt, je seltener und kürzer sie im Vergleich zum ganzen Dokument sind. Treffer in kurzen Feldern werden gegenüber langen Feldern belohnt. Z.B. ist ein Treffer im Titelfeld wichtiger als in einem Abstract oder Volltext, da das Titelfeld im Durchschnitt kürzer ist. Wenn das Relevanzranking nicht benutzt wird, können die Treffer auch nach Erscheinungsjahr absteigend oder nach Titel (alphabetisch, Telefonbuchsartierung DIN 5007-2) sortiert werden.