Volltextsuche

Rastinger, Nina Claudia; ninaclaudia.rastinger@oeaw.ac.at

Im Rahmen einer Volltextsuche wird im gesamten Text eines Dokuments bzw. mehrerer Dokumente innerhalb einer Datenbank gesucht. Im Gegensatz zu anderen Suchstrategien, wie der facettierten Suche, werden eingegebene Termini also nicht mit den Metadaten der jeweiligen Dokumente, sondern mit den von ihnen beinhalteten Wörtern bzw. einem Wort-(formen)-Index abgeglichen und nur bei Übereinstimmung auch gefunden (Gyorodi et al. 2010, S. 736; Beall 2008, S. 438). Dies bedeutet aber auch, dass variierende Schreibungen, wie sie etwa in historischen Texten oftmals vorliegen, ebenso immer mitgedacht werden müssen wie die Tatsache, dass Homonyme der Suchbegriffe ebenfalls gefunden werden: Die Suche nach 'Bank' differenziert nicht zwischen dem Geldinstitut und der Sitzgelegenheit. Ebenso wenig finden Synonyme bzw. andere Ausdrücke für ein gesuchtes Konzept automatisch Eingang in den Suchprozess; diese müssen eigens gewählt und in die Suchleiste eingegeben werden. Infolgedessen hängt der Erfolg einer Volltextsuche nicht nur von der Qualität der jeweiligen Daten ab, sondern immer auch vom individuellen Suchverhalten der Nutzerinnen und Nutzer und der Adäquatheit der von ihnen ausgewählten Keywords (Resch 2019, S. 124; Beall 2008, S. 442).

Um diesen Limitationen entgegenzuwirken und Userinnen und Usern das Bestimmen über den Fein- bzw. Grobheitsgrad ihres Suchprozesses zu erleichtern, werden Volltextsuchen meist über zusätzliche Optionen – wie Wildcards, Boolesche Operatoren oder reguläre Ausdrücke (RegEx) – erweitert: Mit dem RegEx-Suchbefehl /S(a|o)l(a|o)mon*/ etwa werden mehrere Schreibweisen desselben Eigennamens gleichzeitig suchbar. Zudem birgt die Volltextsuche eine Reihe an Vorteilen in sich: So kann das Durchsuchen des wörtlichen Inhalts von Texten – im Sinne eines *Distant Readings* – unter anderem im Hinblick auf die 'Unüberschaubarkeit' geisteswissenschaftlichen Datenmaterials hilfreich sein (Limpinsel 2013, S. 177) und Forschende auf Text(ausschnitt)e stoßen lassen, welche ihnen über eine rein metadatenorientierte Suche eventuell entgangen wären (Kann/Hintersonnleitner 2015, S. 79). Vor diesem Hintergrund macht es Sinn, dass volltextliche Durchsuchbarkeit heutzutage nicht nur im World Wide Web den Standard repräsentiert (Beall 2008, S. 438), sondern zunehmend auch als Zielpunkt für Digitale Editionen und Textsammlungen (vgl. etwa Resch 2019; Kann/Hintersonnleitner 2015) gesetzt wird – denn wie Müller und Hermes-Wladarsch (2017, S. 50) es treffend zusammenfassen: „Erst über durchsuchbare Volltexte potenzieren sich die Möglichkeiten wissenschaftlichen Erkenntnisgewinns."

Literatur:

- Beall, Jeffrey: The Weaknesses of Full-Text Searching. In: The Journal of Academic Librarianship 34: 2008, S. 438–444.
- Gyorodi, Cornelia; Gyorodi, Robert; Pecherle, George; Cornea, George Mihai: Full-Text Search Engine using MySQL. In: International Journal of Computers Communications amp; Control 5: 2010, S. 735–743.
- Kann, Bettina; Hintersonnleitner, Michael: Volltextsuche in historischen Texten. Erfahrungen aus den Projekten der Österreichischen Nationalbibliothek. In: BIBLIOTHEK – Forschung und Praxis 39: 2015, S. 73–79.
- Limpinsel, Marco: Volltextsuche und der philologische Habitus. In: Lesen, Schreiben, Erzählen: kommunikative Kulturtechniken im digitalen Zeitalter. Frankfurt: 2013, S. 171–185.
- Müller, Maria Elisabeth; Hermes-Wladarsch, Maria: Die Digitalisierung der deutschsprachigen Zeitungen des 17. Jahrhunderts – ein Projekt mit Komplexität!. In: Die Zeitung als Medium in der neueren Sprachgeschichte. Korpora – Analyse – Wirkung. Berlin: 2017, S. 39–59.
- Resch, Claudia: Das Wien[n]erische Diarium und seine digitale Erschließung oder „Was die Zeitungsleser vor Geräte haben müssen?“. In: Wiener Geschichtsblätter 74: 2019, S. 115–130.

Software:

elasticsearch, Solr, TextGrid, TUSTEP, TEITOK, Sphinx Search, MySQL

Verweise:

Distant Reading, Facettierte Suche, Historische Korpora, Normalisierung, Transkription

Projekte:

Wien[n]erisches DIGITARIUM, ANNO - AustriaN Newspapers Online , Deutsches Textarchiv, Austrian Baroque Corpus (ABaC:us), Mittelhochdeutsche Begriffsdatenbank (MHDBDB), travel!digital, Uhrfehdebücher der Stadt Basel - Digitale Edition

Themen:

Datenanalyse, Interfaces

Zitiervorschlag:

Rastinger, Nina Claudia. 2021. Volltextsuche. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: https://gams.uni-graz.at/o:konde.211