

V Grundprinzipien der Recherche

In diesem Kapitel werden einige grundlegende Suchfunktionen beim Recherchieren in Informationsressourcen vorgestellt. Obwohl Tipps zum richtigen Suchen und Finden gegeben werden, behandelt es nicht den Gesamtprozess und den Ablauf einer Recherche, die Prinzipien des Auskunftsgesprächs oder den Aufbau einer kontextbezogenen Suchstrategie im Rahmen der Benutzerberatung.

V.1 Historische Ordnungsregeln

Weit mehr als in den heutigen Systemen, die verschiedene Suchzugriffe und eine dynamische Anordnung der Inhalte ermöglichen, stellte in älteren Informationsressourcen die Sortierung von Titeldaten nach klar nachvollziehbaren Regeln eine zentrale Aufgabe dar. Der Suchzugriff in Katalogen, Bibliographien oder Nachschlagewerken aller Art erfolgte zumeist über diese Reihenfolge. Erst seit wenigen Jahren erfolgt die bibliothekarische Erschließung in Deutschland nach dem Regelwerk *Resource Description and Access* (RDA) und damit nach einem internationalen Standard, der speziell für die Recherche in Datenbanksystemen entwickelt wurde. Zuvor galten verbreitet nationale Regelwerke, in Deutschland beispielsweise die *Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken* (RAK-WB; in leicht vereinfachter Form RAK-ÖB für die öffentlichen Bibliotheken), in den USA die *Anglo-American Cataloguing Rules* (AACR2). Gemeinsam ist diesen Regelwerken das Grundprinzip der alphanumerischen Sortierung nach gegebener Wortfolge, teils unter Umgehung des bestimmten oder unbestimmten Artikels am Anfang.

Dieses Grundprinzip der Sortierung nach der gegebenen Wortfolge entspricht den Erwartungen und Suchgewohnheiten der Nutzer, allerdings wurden in früheren Zeiten vielfach andere formale Ordnungsprinzipien verwendet.

Zu den am häufigsten angewandten älteren und teils sehr komplexen Regeln gehörten die *Buchhändlerischen Regeln* und die *Preußischen Instruktionen*. Bibliographische Datensätze aus der Zeit ihrer Anwendung sind inzwischen überwiegend auch in elektronisch zugängliche Kataloge konvertiert und somit fast immer ohne Probleme zu ermitteln. Ungeachtet dessen sollten Informationsspezialisten auch heute noch über die wichtigsten Grundprinzipien informiert sein, nach denen viele Verzeichnisse und Kataloge bis in das letzte Drittel des 20. Jahrhunderts geordnet waren. In besonderer Weise trifft dies auf Personen zu, die in der Bibliothek mit Altdaten arbeiten. Da viele ältere Kataloge und Ver-

zeichnungen heute als Retrodigitalisate online zugänglich sind, kommen auch Benutzer vereinzelt mit den ungewohnten Sortier- und Ordnungsregeln in Kontakt und benötigen dann Unterstützung.

Als Beispiel für die (weitgehende) Anwendung der Buchhändlerischen Regeln kann das *Gesamtverzeichnis des deutschsprachigen Schrifttums 1700–1910 (GV alt)* dienen. Nach den Preußischen Instruktionen (PI), die einige Abweichungen zu den Buchhändlerischen Regeln aufweisen, ist die seit 1925 erscheinende und noch nicht abgeschlossene Druckausgabe des *Gesamtkataloges der Wiegendrucke* (GW s. u. S. 48) geordnet. Dieser verzeichnet und beschreibt die weltweit erschienenen Drucke des 15. Jahrhunderts. Auch das *Gesamtverzeichnis des deutschsprachigen Schrifttums 1911–1965 (GV neu)* ist vorwiegend nach den PI angelegt.

a) Die Buchhändlerischen Regeln

Die buchhändlerischen Ansetzungs- und Ordnungsregeln gelten in den Vorläufern der deutschen Nationalbibliographie (u. a. Heinsius, Kayser, Hinrichs) und teilweise auch im *Deutschen Bücherverzeichnis* (DBV). Partiiell wurden sie auch für das *GV alt* übernommen.

Zu den Besonderheiten der *Buchhändlerischen Regeln* gehören:

- Umlaute werden mit ihrem Grundbuchstaben eingeordnet (ä, ae = a);
- unter Körperschaftsnamen werden keine Eintragungen gemacht;
- die Titel werden in der vorliegenden Form einsortiert, auch bei Übersetzungen.

Die Reihenfolge der Ordnungswörter in einem Sachtitel folgt im Allgemeinen einem grammatischen Prinzip:

1. Erstes unabhängiges Substantiv („Regierendes Substantiv“)
2. dazugehörige Adjektive oder andere abhängige Attribute bzw. „nächst wichtiges Stichwort“.

Artikel, Präpositionen und Konjunktionen werden übergangen. Titel in Satzform oder Titel ohne Substantive werden in gegebener Wortfolge sortiert. Bei Appositionen (zwei Substantive im gleichen Kasus nebeneinander) hat das erste Substantiv Vorrang bei der Sortierung.

Beispiele für die Sortierung (Reihenfolge der Ordnungswörter):

Vorlage: Grundzüge deutscher Literaturgeschichte

Einordnung: *Grundzuge Literaturgeschichte deutscher*

Vorlage: Wir wollen Freiheit, Frieden, Recht
 Einordnung: *Wir wollen Freiheit Frieden Recht*

Vorlage: Welthafen Hamburg
 Einordnung: *Welthafen Hamburg*

Für *anonyme Werke* (dazu zählen auch Werke von mehr als drei Verfassern) gelten zum Teil eigene Regeln:

- Präpositionen (an, von) und Verschmelzungen (*am, vom*) sowie Eigennamen im Genitiv *am Anfang* eines Sachtitels (*Goethes Vermächtnis*) werden erstes Ordnungswort.
- Sind bei anonymen Werken zwei oder mehr Substantive durch Komma oder Konjunktion koordiniert, werden sie bei der Ordnung wie ein zusammengesetztes Wort (Kompositum) behandelt (*Kunst und Handwerk* wird geordnet: *KunstHandwerk*).

Für die Ordnung bei *Verfasserwerken* (bis zu drei Verfasser) gilt:

- Gesamtausgaben (Gesammelte Werke) ordnen vor den Einzelwerken.
- Gemeinschaftlich verfasste Werke ordnen nach den Werken, die vom ersten Verfasser allein stammen.

b) Die Preußischen Instruktionen (PI)

Die Instruktionen für die alphabetischen Kataloge der preußischen Bibliotheken wurden 1899 veröffentlicht, aber erst mit der überarbeiteten Ausgabe von 1908 setzte sich das Regelwerk durch. Es fand auch über Preußen hinaus weite Verbreitung. Trotzdem wurden die PI nicht generell eingeführt, vielmehr existierten an großen Bibliotheken mit historischem Bestand zum Teil abweichende Regeln. So war z. B. an der Bayerischen Staatsbibliothek bis zu Beginn der 1980er Jahre und für Teilkataloge noch länger die *Münchner Katalogordnung* (MKO) in Kraft, die Ähnlichkeiten, aber auch einige Unterschiede zu den PI aufweist.

Die Ordnungsregeln der PI haben u. a. folgende Besonderheiten:

Generell gilt:

- Die Buchstaben i und j werden nicht unterschieden;
- Umlaute werden aufgelöst (ä = ae);
- unter Körperschaftsnamen werden keine Eintragungen gemacht;
- Titel werden mit dem *Originaltitel* eingeordnet, vom vorliegenden (übersetzten) Titel wird verwiesen.

Die Reihenfolge der Ordnungswörter in einem Sachtitel folgt im Allgemeinen einem grammatischen Prinzip:

1. erstes unabhängiges Substantiv
2. dazugehörige Adjektive oder andere abhängige Attribute
3. zweites Substantiv
4. dazugehörige Adjektive oder andere Attribute
5. usw.

Artikel, Präpositionen und Konjunktionen werden übergangen. Titel in Satzform oder Titel ohne Substantive werden in gegebener Wortfolge sortiert. Bei Appositionen hat das *zweite* Substantiv Vorrang.

Beispiele für die Sortierung:

Vorlage: Wissenschaftlicher Berichtsband des deutschen Historikertages

Einordnung: *Berichtsband wissenschaftlicher Historikertag deutschen*

Vorlage: Frühromantische Erzählungen

Einordnung: *Erzaehlungen fruehromantische*

Vorlage: Erzählungen aus Tausendundeiner Nacht

Einordnung: *Alf laila wa-laila, dt.* (und Nebeneintragung oder Verweisung vom vorliegenden Titel)

Vorlage: Welthafen Hamburg

Einordnung: *Hamburg Welthafen*

Für die Ordnung bei *Verfasserwerken* (bis zu drei Verfasser) gilt:

Vor den Einzelwerken eines Verfassers werden Gruppen für folgende Werkgruppen gebildet:

1. *Werke*: Vollständige Werkausgaben;
2. *Teilsammlungen*: Ausgewählte Werke;
3. *Fragmente*;
4. *Werkauszüge*: Sammlungen mit Auszügen aus Werken.

Im Anschluss an diese Gruppen folgen die Einzelwerke, die mit ihrem *Originalsachtitel* eingeordnet werden.

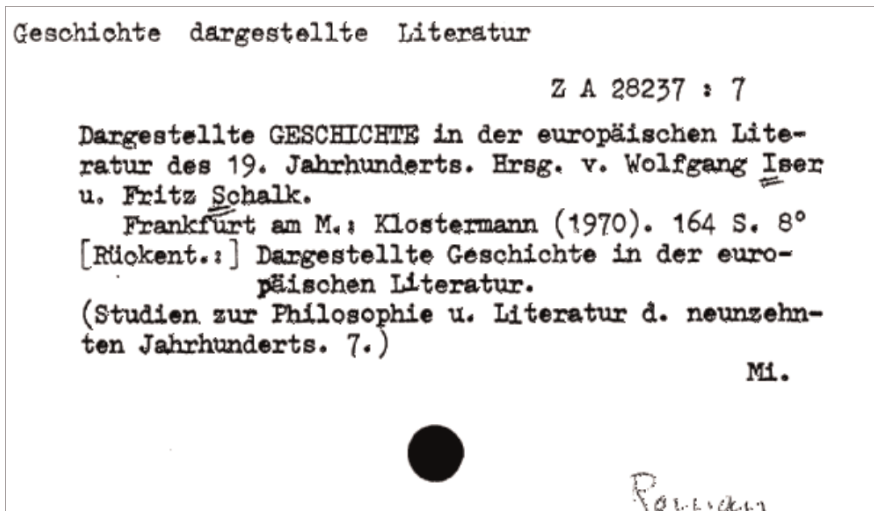


Abb. 5: Ordnung eines Sachtitels nach den PI (Image-Katalog der SUB Göttingen); weitere Eintragungen (Iser, Schalk) werden durch den Unterstrich kenntlich gemacht

Gedruckte Verzeichnisse und digitalisierte Kartenkataloge, die nach historischen Regeln geordnet sind, enthalten häufig Hinweise und Anleitungen zur Ansetzung von Namen und zur Ordnung der Sachtitel, die im Zweifelsfall zu beachten sind.

V.2 Retrieval in Datenbanken

In einer Online-Umgebung besteht der Suchvorgang nach den *Functional Requirements for Bibliographic Records* (FRBR) aus folgenden Schritten:

1. Finden (Find): den Suchbedarf formulieren und Treffer erhalten;
2. Identifizieren (Identify): die gefundene Treffermenge sichten und prüfen, ob relevante Suchergebnisse ermittelt wurden; nötigenfalls Suche in geänderter Form wiederholen;
3. Auswählen (Select): aus den relevanten Treffern diejenigen auswählen, die letztendlich gewünscht werden, z.B. eine bestimmte Auflage oder eine Übersetzung in eine bestimmte Sprache;
4. Zugang erhalten (Obtain): die benötigten Dokumente erhalten bzw. bestellen.

Das folgende Kapitel dient der allgemeinen Erläuterung einiger grundlegender Retrievalfunktionen in elektronischen Informationsmitteln. Trotz allgemeiner Konventionen können Suchfunktionen in den einzelnen Ressourcen zum Teil auch sehr unterschiedlich gestaltet sein. Während das Verhalten der Benutzer geprägt ist von der Recherche mit allgemeinen Suchmaschinen, sollten Spezialisten immer die erläuternden Texte zur Suche in Informationsmitteln beachten, da sie oft weiterführende Recherchemöglichkeiten bieten (z. B. mehr Suchkriterien als im Drill-Down-Menü angezeigt werden, Suchmöglichkeiten mit Fachsystematiken etc.) oder auf Besonderheiten der Recherche hinweisen. Zum Beispiel werden bei einigen wenigen Aufsatzdatenbanken Treffer eines Autors zum Teil nur dann gefunden, wenn als Vorname ausschließlich der erste Buchstabe des ersten Vornamens eingegeben wird. Grund dafür ist, dass gerade in amerikanischen Datenbanken die Autoren zeitweise nur so erfasst wurden. Der Vorname muss also zwingend abgekürzt werden. Eine Besonderheit des *Web of Science* besteht beispielsweise darin, dass die Titel von Aufsätzen immer ins Englische übertragen werden und vielfach unter dem jeweiligen Originaltitel nicht recherchiert werden können.

a) Grundbegriffe

Einfache Suche – Feldspezifische Suche – Indexsuche

Bibliographische Daten sind in Felder gegliedert. Typische Felder sind: Autor, Sachtitel, Erscheinungsort, Erscheinungsjahr, Verlag, Schlagwörter. Im Detail ist die Titelaufnahme oft weit ausführlicher strukturiert und codiert und erlaubt dementsprechend auch wesentlich präzisere Recherchen. Die Suchoberflächen, die z. B. für Bibliothekskataloge (s. u. S. 68) konfiguriert werden, haben in der Regel zwei Oberflächen: eine Einfache Suche, die google-ähnlich in einer Zeile ausgeführt wird, sowie eine Erweiterte Suche, die feldbezogen und mit Booleschen Operatoren (s. u.) ausgeführt wird.

Die *Einfache Suche* (auch Freitextsuche, Basic Search oder Simple Search) ist eine umfassende Suche, die auf mehrere indexierte Felder zugreift, wobei dies nicht alle Felder umfassen muss, die in der Expertensuche gezielt angesteuert werden können. Sie ist empfehlenswert, wenn unspezifisch gesucht wird oder ungenaue Daten vorliegen. So können z. B. bei einer Suche nach dem Wort Energiekrise Treffer gefunden werden, bei denen der Begriff im Titel, im Zusatz oder als Schlagwort vorkommt. Mit der einfachen Stichwortsuche wird teilweise auch der Text eingescannter Zusatzinformationen erreicht. Wenn eine angezeigt-

te Titelaufnahme den Suchbegriff nicht enthält, kann er z. B. Bestandteil des Inhaltsverzeichnisses sein, das über einen Link verknüpft ist.

Die *Erweiterte Suche* (auch Expertensuche, Feldspezifische Suche, Advanced Search oder Guided Search) ermöglicht die Einschränkung der Suche auf genau definierte Kriterien: alle Werke eines Autors; alle Dokumente, in denen ein Stichwort im Sachtitel vorkommt; alle Dokumente, die ein Schlagwort aufweisen, etc. In vielen Fällen ist bei der feldspezifischen Suche optional auch die Suche im *Index* (Register) des Feldes möglich. Der gesuchte Begriff (der genaue Name des Autors, das Titelstichwort, das korrekte Schlagwort) kann dann aus einer angebotenen Liste ausgewählt und in das Suchfeld übernommen werden. Diese Funktion entspricht einem Browsing, sie reduziert Fehlerquellen und ist unbedingt empfehlenswert. Gerade bei Datenbanken aus dem angloamerikanischen Bereich sind Verweisungen von weiteren Namensformen einer Person oder einer Körperschaft häufig nur über den Index zu finden.

Stichwort – Schlagwort – Thesaurus – Deskriptor

Die meisten bibliothekarischen Suchsysteme unterscheiden zwischen Stichwortsuche und Schlagwortsuche. Ein *Stichwort* (keyword) ist im engeren Sinn ein Wort aus dem Titel. Die Stichwortsuche kann sich jedoch auch auf weitere Felder der Ressourcenbeschreibung beziehen, z. B. auf den Titelzusatz, die Verantwortlichkeitsbeziehungen (Autor, Körperschaft), die Verlagsangabe; letztendlich kann der gesamte Text der bibliographischen Beschreibung oder auch ein Abstract oder der Volltext des Werkes für die Stichwortsuche angeboten werden.

Ein *Schlagwort* (subject heading) ist unbedingt zu unterscheiden von einem im bibliographischen Text vorkommenden Stichwort. Schlagwörter werden intellektuell erstellt; ein Schlagwort oder mehrere Schlagwörter (Schlagwortfolge) beschreiben möglichst genau den Inhalt eines Dokuments. Bei der bibliothekarischen Erschließung stammen die Begriffe aus einem kontrollierten Vokabular, das in der *Gemeinsamen Normdatei* (GND) hinterlegt und dort aktuell gehalten wird. Die richtigen Begriffe – in der Regel in der Singularform – können in Suchsystemen oft aus einer Liste (Index) ausgewählt werden.

In einem *Thesaurus* werden zueinander in Beziehung stehende Begriffe (Deskriptoren), Synonyme und verwandte Begriffe aus einem Sachgebiet hierarchisch (d. h. in Form von Ober- und Unterbegriffen) geordnet und strukturiert angezeigt.

Wenn Literatur zu einem Thema gesucht wird, dann ist eine Stichwortsuche im Titel und im Titelzusatz nicht ausreichend. Es sollte zusätzlich die Schlagwortsuche benutzt werden. Auf diese Weise können z. B. auch fremdsprachige Werke gefunden werden. Zu beachten ist allerdings, dass die Schlagwortvergabe

nach einheitlichen Regeln im deutschsprachigen Raum erst seit den 1980er Jahren einsetzte; gerade bei älteren Titeln fehlen sie häufig vollständig. Aus diesem Grund wird für die Schlagwortsuche oft das Titelfeld ebenfalls berücksichtigt, um so zumindest diejenigen Treffer zu erhalten, bei denen ein gesuchtes Schlagwort im Titel vorkommt. Da Erschließungsdaten international genutzt und ausgetauscht werden, können bei der Übernahme von Fremddaten in den eigenen Katalog auch Schlagwörter übernommen werden, die nicht unbedingt den eigenen Regeln entsprechen, z. B. aus einer anderen Sprache stammen. Vielfach werden sie dann mit deutschen Schlagwörtern ergänzt.

Die Bildung von Schlagwörtern ist in Deutschland in den *Regeln für den Schlagwortkatalog* (RSWK) festgelegt. Es gilt das Prinzip des spezifischen Schlagwortes, d. h. einem engeren Begriff (Buchmalerei) wird der Vorzug vor einem weiten Begriff (Malerei) gegeben. Neben Sachschlagwörtern werden auch Personenschlagwörter, geographische Schlagwörter und Zeitschlagwörter (für einen Zeitpunkt oder Zeitraum) vergeben. Auch Formalangaben werden berücksichtigt (z. B. Angaben zur physischen Form der Veröffentlichung).

Beispiel einer Schlagwortfolge:

Das Werk *Bauhaus, Modernism and the Illustrated Book* hat folgende Schlagwortfolge:

Bauhaus ; Illustration ; Druckwerk ; Geschichte

Die Einzelbegriffe sind ebenfalls suchfähig. Es kann also auch mit den Schlagwörtern *Illustration* oder *Druckwerk* gesucht werden.

Klassifikatorische Suche

Bei der sachlichen Erschließung von Literatur ist neben der Schlagwortvergabe auch die Systematisierung des Inhalts nach einem Klassifikationssystem der Wissenschaften verbreitet. Das hierarchische System besteht aus Haupt- und Unterklassen, die sich in immer speziellere Systemstellen weiter untergliedern lassen.

Der Vorteil der klassifikatorischen Erschließung besteht darin, dass sowohl sehr themenspezifisch als auch im größeren Zusammenhang eines ganzen Fachgebietes Literatur gesucht werden kann; auch die Unabhängigkeit von Einzelsprachen und sprachlichen Zweifelsfällen bildet einen Vorteil. Neben einer Gesamtübersicht, die eine Orientierung im System ermöglicht, gehört als Hilfsmittel zu einer Systematik zumeist ein Register mit Begriffen, die auf die entsprechende Systemstelle bzw. ihre Bezeichnung (Notation) verweisen. Die Notation besteht aus einem hierarchisch geordneten System nur aus Zahlen oder aus Zahlen und Buchstaben. Sie wird bei Bibliotheken mit Freihandaufstel-

lung häufig als Teil der Signatur verwendet. Als Klassifikationssysteme sind im deutschen Bibliothekswesen die auch international gebräuchliche *Dewey Decimal Classification* (DDC) und die *Regensburger Verbundklassifikation* (RVK) von besonderer Bedeutung. Die Suche mit der Notation aus einem Klassifikationssystem ist in vielen Suchsystemen möglich, z. B. in der *Deutschen Nationalbibliografie* (DDC) und in Verbundkatalogen (RVK und weitere dort verwendete Systeme). Von der Website der RVK aus kann hierarchisch navigiert und an der gewünschten Stelle in Verbundkatalogen des deutschsprachigen Bereichs gewechselt werden (BVB, GBV, SWB, OBV und swisscovery).

▼ **AA Allgemeines**

▼ **AA Bibliografien der Bibliografien, Universalbibliografien, Bibliothekskataloge, Nationalbibliografien**

AA 09900 Bibliografische Zeitschriften

▼ **AA 10000 - AA 19900** Bibliografien der Bibliografien

AA 10000 International, Allgemeines

AA 10100 Antike Welt

AA 10300 Ländergruppen/Entwicklungsländer

▶ **AA 10400** Westliche Welt, Abendland, Europäische Welt

▼ **AA 10600** Europa

AA 10640 Europäische Gemeinschaften, Europäische Union

▼ **AA 10700** Mitteleuropa

AA 10720 Alpenländer

AA 10760 Österreich-Ungarische Monarchie

AA 10800 Deutsches Sprachgebiet

Abb. 6: Ausschnitt aus der Regensburger Verbundklassifikation online (RVK)

Normdaten

Die Ansetzung von Personennamen, Körperschaftsnamen und Schlagwörtern erfolgt im Rahmen der Katalogisierung über kooperativ geführte Normdateien (authority files). Im deutschsprachigen Bereich wurden die bis 2012 getrennt geführten Datenbanken für Personen, Körperschaften und Schlagwörter zur *Gemeinsamen Normdatei* (GND) mit mehr als neun Millionen Datensätzen zusammengeführt. Auch für Sachtitel gibt es genormte Formen (Einheitssachtitel, uniform title). So können bei häufig wechselnden Werktiteln und bei Übersetzungen (Nibelungenlied – Lied der Nibelungen – Siegfriedsaga – Song of the Nibelungs) Werke zusammengeführt werden. Besonders erforderlich sind Einheitssachtitel auch bei Musikalien. Details zur Festlegung von Namensfor-

men und zusätzlich erfasste Daten wie z. B. Lebensdaten oder Beziehungen zwischen Personen und Körperschaften untereinander werden ebenfalls in der *GND* verankert. Auf internationaler Ebene wird mit der *Virtual International Authority File* (VIAF) eine Konkordanz für unterschiedliche Ansetzungsformen geführt. Denn auch das internationale Regelwerk RDA lässt weiterhin nationale und nationalsprachliche Ansetzungsformen zu, auch wenn die Grundsätze für die Normierung generell geregelt sind.

Bei der Formalkatalogisierung garantieren Normdaten die eindeutige Benennung von Personen, Körperschaften und Werken; bei der Inhaltserschließung stellen sie ein einheitlich verwendetes kontrolliertes Vokabular bereit. Somit dienen Normdaten der Identifizierung, der Referenzierung und der Vernetzung von Entitäten sowie der Auffindbarkeit und Nachnutzbarkeit von Daten. Die Erschließung mit Normdaten ist ein besonderes Charakteristikum für Bibliothekskataloge, die sich dadurch von anderen Suchinstrumenten, vor allem auch von allgemeinen Suchmaschinen, erheblich unterscheiden. Wenn bei der Recherche in einem Verzeichnis die eingegebenen Suchbegriffe nicht der genormten Ansetzung entsprechen, können sie trotzdem gefunden werden, da in den Normdaten zahlreiche weitere Formen hinterlegt sind.

Beispiel:

Für den Namen des Dichters Anton Pawlowitsch Tschechow (gängige Schreibweise in deutschsprachigen Nachschlagewerken) sind in der *GND* rund 90 Namensformen hinterlegt, die teilweise aus anderen Schriften transkribiert oder transliteriert sind, darunter:

Čechonte, Antoša

Čechow, Anton P.

Çexov, Anton

Chekhov, Anton Pavlovich

Chì-kò-fu

Czechow, Antoni

Tchékhov, Anton Pavlovitch

Von allen Formen und Schreibweisen wird auf den „bevorzugten Namen“

Čechov, Anton Pavlovič verwiesen.

Normdaten dienen auch dem Austausch und der Vernetzung über den Bereich der Bibliothekswelt hinaus. Sie erleichtern die Herstellung von Zusammenhängen. Zunehmend werden die Normdaten der *GND* auch in anderen Bereichen verwendet, z. B. im Archiv- und Museumsbereich, um so eine umfassende Kontextualisierung zusammengehöriger Inhalte zu ermöglichen.

b) Organisation der Suchanfrage

Welche Suchmethoden ausgeführt werden, ist sowohl vom konkreten Einzelfall als auch von den Funktionen abhängig, die eine Datenbank bzw. eine Suchmaschine anbietet. Wenn mehrere Suchbegriffe mit Booleschen Operatoren verknüpft werden, ist am ehesten mit einem exakten Ergebnis – *Exact Match* – zu rechnen. Die andere Möglichkeit – *Best Match* – wird erreicht, wenn ein Rankingverfahren zum Einsatz kommt, das neben exakten Treffern auch ähnliche Ergebnisse auflistet. Die Treffermenge wird dann größer und vielfältiger; sie kann jedoch auch ungenauer sein. Informationsexperten können Nutzer beraten, welche Suchstrategie im konkreten Fall geeignet ist. Im Folgenden werden einige für die Suche generell beachtenswerte Punkte aufgeführt.

Boolesche Operatoren und Abstandsoperatoren

Für die Eingabe mehrerer Begriffe in einer Zeile oder für die Verknüpfung mehrerer Felder in einem Suchvorgang werden die nach dem englischen Mathematiker George Boole benannten *Booleschen Operatoren* „UND“, „ODER“, „NICHT“ verwendet. Sie werden in der Regel über ein Menü ausgewählt, in seltenen Fällen müssen sie von Hand eingegeben werden. In diesem Fall ist es notwendig, die in der Datenbank vorgegebene Schreibweise – groß oder klein, deutsch oder englisch, Zeichen anstatt der Wörter – zu beachten. Eine Suche mit dem UND-Operator ermittelt Treffer, in denen beide Begriffe vorkommen. Mit einer ODER-Suche wird die Treffermenge erweitert, sie zeigt alle Datensätze an, in denen entweder der eine oder der andere oder beide Begriffe vorkommen. NICHT schließt Datensätze mit dem entsprechenden Suchbegriff explizit aus. Es ist eine allgemeine Konvention in Suchmaschinen, dass mehrere eingegebene Begriffe automatisch mit UND verknüpft werden, sofern keine Option gewählt wird. Für komplexe Recherchen können die Suchanfragen mit Hilfe von Klammern logisch verknüpft werden, um so die Reihenfolge der Suchschritte festzulegen.

Beispiel:

(analphabetismus OR illiteracy) NOT (kinder OR children)

Gefunden werden Aufsätze in deutscher oder englischer Sprache zum Analphabetismus, jedoch nicht bei Kindern.

Mit *Abstandsoperatoren* kann der genaue oder der maximale Abstand zwischen zwei Suchbegriffen festgelegt werden. Durch einen engen Abstandsoperator kann man sicherstellen, dass die beiden Suchbegriffe in einem engen inhaltli-

chen Zusammenhang zueinander stehen; in Einzelfällen kommen auch *Vergleichsoperatoren* zum Einsatz.

Schreibweise und Stoppwörter

Bei den meisten Suchsystemen kann heute die vorliegende Schreibweise für Umlaute, Sonderzeichen und Zahlen etc. beibehalten werden. Diakritische Zeichen werden bei der Suche nicht eingegeben. Umlaute wie ä, ö usw. werden in dieser Form gefunden. In angloamerikanischen Datenbanken empfiehlt es sich, eine Testsuche durchzuführen. Dort war es lange Zeit üblich, Umlaute obligatorisch auf ihre Grundbuchstaben zurückzuführen. Inzwischen kann in der Regel in Datenbanken auch mit dem Umlaut gesucht werden, vielfach jedoch nicht mit der aufgelösten Form. Bei Sonderzeichen wie ß ist eine Testsuche ebenfalls empfehlenswert.

Beispiel aus dem Katalog der Library of Congress:

Eine Suche nach dem Wort *münchen* im Titel ergibt rund 8 000 Treffer. Die gleichen Treffer werden angezeigt, wenn mit dem Wort *munchen* gesucht wird. Wird jedoch mit der aufgelösten Form *muenchen* gesucht, so werden nur die Treffer gefunden, bei denen die Schreibweise *muenchen* tatsächlich vorliegt.

In einigen wenigen Suchsystemen sind sogenannte Stoppwörtern als Begriffe von der Suche ausgeschlossen, zum Teil werden diese Begriffe in einer Liste angezeigt. Dazu gehören vor allem häufig vorkommende Wörter wie Artikel und Pronomen.

Komposita, Wortverbindungen, Groß- und Kleinschreibung

Mit Komposita und Wortverbindungen mit Bindestrich gehen Suchsysteme uneinheitlich um. In modernen Katalogdatenbanken werden diese Wörter zumeist doppelt indexiert, so dass sie sowohl als ein Wort oder als zwei Wörter gesucht werden können. Grundsätzlich empfiehlt sich bei schwankenden Schreibweisen eine Testsuche, da die Regelungen dazu unterschiedlich sind.

Es ist gängige Konvention, dass Suchbegriffe mit Klein- oder Großbuchstaben geschrieben werden können. Allerdings kann bei einigen Suchsystemen auch ausdrücklich mit Großbuchstaben gesucht werden. Die Suche wird dann anders ausgeführt. Es werden nur Treffer mit Großbuchstaben angezeigt oder diese Treffer werden bevorzugt angezeigt. Die Funktion kann nützlich sein, wenn Folgen von Initialen gesucht werden.

Beispiel:

man findet die Treffer „man“ und „MAN“;

MAN findet nur die Treffer „MAN“ (zur Fahrzeugfirma).

Phrasensuche, Stringsuche, exakte Titelsuche

Um die Suche präziser ausführen zu können, ist vielfach die Phrasensuche (exakte Suche) hilfreich. Hierbei wird nach einer festen Abfolge von Wörtern (Phrase) gesucht, die Treffermenge kann damit gegenüber der voreingestellten Oder-Verknüpfung der Suchbegriffe eingeschränkt werden. Die Phrasensuche wird in vielen Systemen durch Einschluss in Anführungszeichen ausgedrückt, z. B. „manchester united“. Der Begriff Stringsuche bezeichnet noch allgemeiner die Suche nach einer feststehenden Folge von Zeichen oder Wörtern.

Während Phrasen- und Stringsuchen an beliebigen Stellen eines Suchtextes starten können, muss die exakte Titelsuche vielfach mit dem Beginn des Titels oder mit dem kompletten Titel ausgeführt werden. Dies kann etwa bei der Suche nach bestimmten Zeitschriftentiteln mit häufig vorkommenden Wörtern, z. B. bei der Wochenzeitung „Die Zeit“, sehr hilfreich sein.

Trunkierung, unscharfe Suche

Viele Recherchesysteme bieten die Möglichkeit, einzelne Suchbegriffe entweder exakt oder unscharf zu suchen. Einige haben bereits eine Voreinstellung, die Flexionen oder regelmäßige Pluralbildungen von Suchbegriffen mit einbeziehen, andere suchen auch phonetisch, d. h. es werden auch Begriffe angezeigt, die zwar anders geschrieben, jedoch ähnlich gesprochen werden. Ist dies nicht erwünscht, kann der Begriff exakt als „Phrase“ gesucht werden. Die gängigste Methode, verschiedene Wortformen einzuschließen, ist die Trunkierung (Wortstammsuche). Dabei werden Bestandteile eines Wortes und ein spezielles Trunkierungszeichen eingegeben.

Beispiel:

*recycl** findet „recycle“ und „recycling“.

Typischerweise wird die Rechtstrunkierung angeboten; hierbei ersetzt der Platzhalter die rechts vom Wort stehende Wortendung. Von Linkstrunkierung spricht man, wenn der linke Teil eines Begriffs bei der Suche offen bleibt (**epik* findet „heldenepik“ und „versepik“. Als Trunkierungszeichen sind vor allem die Zeichen * und ? gebräuchlich.

Um Schreibvarianten im Inneren eines Begriffes zu suchen, können Platzhalter (wildcard; joker) benutzt werden. Diese Form der Sucheingabe wird auch

als Maskierung oder Binnentrunkierung bezeichnet. Dabei ist darauf zu achten, ob das entsprechende Zeichen (z. B. #) jeweils nur einen einzelnen Buchstaben ersetzt oder ob es auch für mehrere Zeichen stehen kann (*mi#rofilm* findet „mikrofilm“ und „microfilm“).

Semantische Suche

Einzelne Suchsysteme verwenden Verfahren, die über eine Suche mit Stichwörtern oder Schlagwörtern hinausgehen und die Suchanfrage stattdessen auf ihre Bedeutung hin analysieren. Dazu wird die Anfrage im Hintergrund mit Fachthesauri abgeglichen, die wiederum aufgebaut werden aus Ontologien, Taxonomien, Klassifikationen und Fach- oder Sprachwörterbüchern und in einem „Semantischen Netz“ die Beziehungsstrukturen von Begriffen bzw. deren Konzept auswerten. Damit können Suchergebnisse erzielt werden, die den sachlichen Kern der Anfrage treffen, ihren Sinn verstehen, ohne dass bei der Suchanfrage tatsächlich die „richtigen“ Begriffe verwendet werden müssen. Die semantische Suche führt in der Regel zu einer deutlichen Erhöhung der Trefferzahlen, da für die Suche neben dem eingegebenen Begriff weitere Suchterme verwendet werden; die Relevanz der Treffer muss im Einzelfall geprüft werden. Auch für die inhaltliche Gruppierung von Treffermengen und deren visuelle Darstellung kann die semantische Suche eingesetzt werden.

Originalschriftliche Suche

Es ist zunehmend möglich, Datensätze in nichtlateinischen Schriften auch originalschriftlich zu suchen. Dies reduziert die Fehlerquelle, die entsteht, wenn ein Titel nach unterschiedlichen Umschriftsystemen transkribiert oder transliteriert wird. Die originalschriftliche Katalogisierung und Suche setzt komplexere technische Einstellungen voraus. Realisiert ist diese Funktion z. B. in Bibliothekssystemen, in Fachinformationsdiensten, in Regionalportalen für entsprechende Sprachen und Kulturen sowie in einzelnen Bibliothekskatalogen mit Beständen in nichtlateinischen Schriften. Für die Eingabe der originalschriftlichen Suchbegriffe bieten manche Suchsysteme virtuelle Tastaturen mit den entsprechenden Belegungen an.

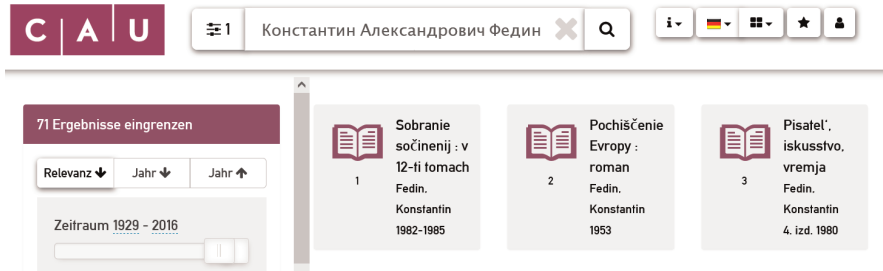


Abb. 7: Originalschriftliche Suche im Discovery-System der UB Kiel

Retrievalssprachen

Zusätzlich zu einer über das Menü gesteuerten Abfrage verfügen vor allem Datenbanken im naturwissenschaftlich-technischen Bereich über teils proprietäre Abfragesprachen (Retrievalssprachen, Kommandosprachen), die komplexe Strategien mit Hilfe einer detaillierten Suchsyntax ermöglichen. Für die indexierten Felder werden Codes verwendet, zur Verknüpfung dienen die Booleschen Operatoren oder deskriptive Zeichen. Auch Fachthesauri können einbezogen werden. Einzelne Suchschritte können so editiert und neu kombiniert, die Ergebnisse analysiert und in Tabellenform ausgegeben werden.

Beispiel aus der Retrievalssprache Messenger:

S PLASMA/TI,CT

Eine Suche (S) nach dem Begriff „plasma“ wird im Feld Titel (TI) und im Thesaurus (CT, Controlled Term) ausgeführt.

Nicht-textuelle Suche

Fast alle Suchzugriffe in Informationsressourcen, Suchmaschinen oder vergleichbaren Diensten basieren auf der Eingabe von sprachlichen, textbasierten Inhalten (meist Stich- oder Schlagwörter), auch numerische oder alphanumerische Kombinationen (Erscheinungsjahr oder Notation einer Klassifikation) zählen zu dieser textbasierten Suche. Zunehmend werden jedoch auch Verfahren für nicht-textbasierte Rechercheverfahren entwickelt, insbesondere um so alternative Zugriffe auf Nicht-Text-Materialien (Bilder, Audios, Karten) zu ermöglichen. Wichtige Verfahren hierbei sind vor allem die *georeferenzierte Suche*, die

Bildähnlichkeitssuche und die *Melodiesuche* für Musikalien und auditive Materialien. Für die Entwicklung innovativer Infrastrukturen, Werkzeuge und Dienstleistungen für das Auffinden, Durchsuchen und Zitieren von Videos, Forschungsdaten, Grafiken und Open Educational Resources (OER) wurde 2012 an der TIB Hannover das Lab Nicht-Textuelle Materialien (NTM) gegründet.

Georeferenzierte Suche

Die georeferenzierte Suche ermöglicht für die Recherche nach entsprechend referenzierten Materialien die Eingabe von Suchkriterien in Form von normierten Geokoordinaten (*Geographic Coordinate System*, GCS). Die Vergabe von geographischen Informationen (Geotags) bei der Erschließung von Dokumenten kann von Hand eingegeben werden oder auch automatisiert erfolgen (z. B. bei Aufnahmen von Digitalkameras über GPS-Kennung). Die Auswahl von Geokoordinaten für die Recherche kann über die direkte Eingabe eines gewünschten geographischen Bezugspunktes erfolgen, über die verbale Auswahl eines Geographikums aus einer Liste mit entsprechend referenzierten Einträgen oder auch über das Anklicken eines Punktes bzw. das Markieren einer Fläche auf einer digitalen Landkarte. Besonders häufig zum Einsatz kommt die georeferenzierte Suche vor allem bei Informationsressourcen im Bereich von Landkarten und Regionalportalen. Aber auch klassische Fachbibliographien oder Verbundkataloge nutzen bereits raumbezogene Informationen für die Recherche und die Anzeige ihrer klassisch-textbasierten Suchergebnisse. So kann beispielsweise in der *Georeferenzierten Online-Bibliographie Areallinguistik* (GOBA) über das sprachgeographische Informationssystem *REDE SprachGIS* anhand von Karten nach Literatur zu einem bestimmten Ort oder Gebiet gesucht werden. Wird im *WorldCat* der eigene Standort eingegeben oder durch die automatische Zuordnung zu einem Standort über die IP-Adresse vorgenommen, so erfolgt die Anzeige der weltweiten Besitznachweise gesuchter Ausgaben in einer nach geographischen Kriterien sortierten Liste. Die nächstgelegene Bibliothek mit dem entsprechenden Bestandsnachweis steht am Beginn dieser Liste, das Ende der Liste bildet die Bibliothek, die am weitesten entfernt vom eigenen Standort liegt.

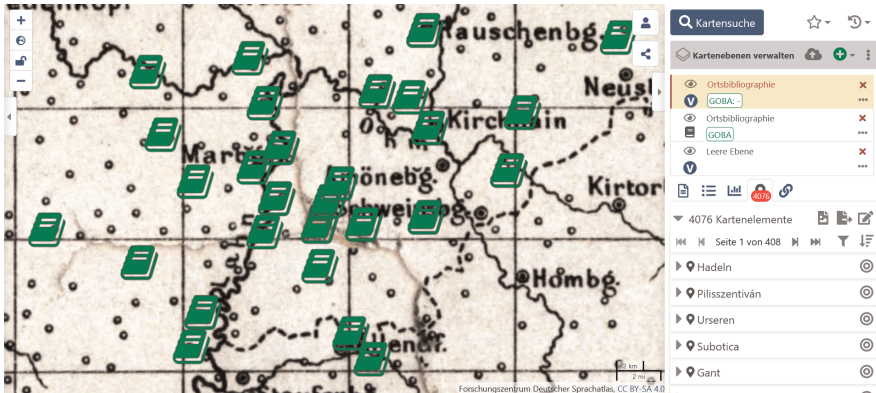


Abb. 8: Trefferanzeige ortsbezogener Fachliteratur in der Georeferenzierten Online-Bibliographie Areallinguistik (GOBA)

Bildähnlichkeitssuche

Um Informationsressourcen oder das Internet nach passenden Bildern zu durchsuchen, können sowohl verbale Elemente genutzt werden, die entweder Erschließungsdaten (Bilddatenbanken) oder den begleitenden Text von Bildern (Suchmaschinen) nutzen. Neben diesen Verfahren gibt es jedoch auch die Möglichkeit, eine Bildähnlichkeitssuche zu verwenden. Hierfür wird ein ausgewähltes oder eigens hochgeladenes Bild als Ausgangspunkt für den Vergleich mit den Bildern im Suchraum verwendet. Typischerweise nutzen solche Bildähnlichkeitssuchen Kriterien wie Texturen, markante Merkmale, Kontrast- und Farbverteilung sowie Bildgrößen. Für die tatsächliche Suche genutzt werden dann jedoch nicht die Bilder, sondern visuelle Deskriptoren, in denen der Bildinhalt niedergelegt ist. Entsprechende Angebote gibt es sowohl im Bibliotheksbereich (Bildähnlichkeitssuche der Bayerischen Staatsbibliothek) als auch bei kommerziellen Suchdiensten (Google Bilder etc.).

Weitere Verfahren der Bilderschließung und der Bildrecherche bieten auch die heutigen Möglichkeiten der Bildinhaltsanalyse durch *Deep Learning* und *Künstliche Intelligenz* (KI). Künstliche neuronale Netzwerke (*Artificial Neural Network*, ANN) werden hierbei mit großen Mengen erschlossener Bilder trainiert und erlernen in diesem Prozess eigenständig die Erkennung von charakteristischen Strukturen. Mit Hilfe dieses Verfahrens kann bei der Recherche nach Bildern z. B. nach festgelegten Bildinhalten gesucht werden (Bilder ohne Personen, mit einer Person, zwei Personen oder einer beliebigen Anzahl; mit Männern, Frauen, Kindern oder älteren Menschen, mit Lachen, ohne Lachen etc.), die Suche kann sich auf automatisch vergebene Stichwörter beziehen, aber auch eine

Suche mit vorgegebenen Vergleichsbildern ist möglich. Bildanalyse auf der Grundlage von Künstlicher Intelligenz kommt in vielen Bereichen, etwa der Medizin, der Versicherungswirtschaft oder des E-Commerce zum Einsatz, doch auch Bibliotheken wie die TIB entwickeln Verfahren zur Geolokalisierung von Fotos mithilfe künstlicher neuronaler Netze (Projekt *Geolocation Estimation*).

Melodiesuche

Auch nach Musiknoten und Audiodokumenten kann nicht nur mit Hilfe von textuellen Metadaten (Titeldaten, Schlagwörtern etc.) gesucht werden; wie bei Bildern kann auch hier die Suche durch eingegebene Ton- oder Notenfolgen erfolgen (Music Information Retrieval).

Für die Suche nach Audiodokumenten werden diese zunächst in Bezug auf signifikante Merkmale hinsichtlich der Tonhöhe und des Rhythmus analysiert, die dann für den Vergleich mit eingegebenen Melodien zur Verfügung stehen. Anwendungen für diese Art der Melodiesuche gibt es sowohl im kommerziellen Bereich (*midomi*) als auch im Bereich von Nachschlagewerken (*musipedia*).

In vergleichbarer Weise kann auch nach Noten recherchiert werden. Das Notenmaterial gedruckter Noten muss hierfür zunächst mit Hilfe einer automatischen Notenerkennung (Optical Music Recognition) in maschinenlesbare Form gebracht werden. Häufig wird dieser Prozess nur auf den Beginn von Musikstücken angewendet (Musikincipits). Nach diesem Notenmaterial kann dann – wiederum über eingegebene Noten – recherchiert werden. Die Eingabe der Noten erfolgt in der Regel über eine virtuelle Klaviatur am Bildschirm (vgl. *RISM Opac*, s. u. S. 613). Ein Rankingverfahren sortiert die Musikalien, Noten, die der eingegebenen Notenfolge am ähnlichsten sind, erscheinen oben in der Trefferliste, Noten mit größeren Abweichungen weiter unten.

c) Datenformate

Technische Datenformate, mit denen in Bibliotheken vorzugsweise gearbeitet wird, zeichnen sich vor allem durch einen hohen Grad der Normierung aus. Sie sind dafür entwickelt, den nationalen und internationalen Datenaustausch so gut wie möglich zu unterstützen, Nachnutzungen in anderen Kontexten zu ermöglichen und eindeutige Rechercheergebnisse zu gewährleisten. Die technischen Datenformate, in denen die Daten verarbeitet werden, sind zu unterscheiden von Anzeigeformen und -konventionen, die es heute in sehr großer, kaum noch zu überschauender Zahl gibt (Zitierstile). Mit ISBD, MARC 21 und Dublin

Core werden im Folgenden drei Formate wegen ihrer weiten Verbreitung kurz vorgestellt.

3 von 13 Ihre Aktion: suchen [und] (Suche über alles [ALL]) brecht bertolt kugli 3 v

Felder \ MARC21 (SWB-PPN) \ MARC21 (K10plus-PPN) \ ISBD \ Citavi, Referenzmanager (RIS) \ Endnote Tagged Format \ BibTex-Format \ RDF-Format

K10plusPPN: 1612767222
SWB-ID: 279701446

Titel: Der kaukasische Kreidekreis / Bertold Brecht. Mit einem Kommentar von Ana Kugli. - Orig.-Ausg., 1. Aufl., [Nachdr.]

Autorin/Autor: Brecht, Bertolt, 1898-1956

Beteiligt: Kugli, Ana, 1975- [Komment.]

Erschienen: Frankfurt am Main : Suhrkamp, 2007

Umfang: 188 S.

Schriftenreihe: Suhrkamp-BasisBibliothek ; 42 : Text und Kommentar

ISBN: 3-518-18842-9

RVK-Notation: GM 2658 → Ähnliche Literatur

Schlagwortfolge: *Brecht, Bertolt / Der kaukasische Kreidekreis → Zum Register

Mehr zum Titel: Inhaltsverzeichnis

Buchhandel?

Google Buchsuche

Abb. 9: Standardanzeige einer Titelaufnahme im Verbundkatalog des SWB

ISBD-Format

Das insbesondere in Kartenkatalogen viele Jahrzehnte dominierende Anzeigeformat für bibliographische Daten ist die *International Standard Bibliographic Description* (ISBD). Das Format legt für verschiedene Medienarten wie Bücher und Zeitschriften die spezifische Form der bibliographischen Beschreibung fest. Die Daten sind in kompakter Form und als relativ zusammenhängender Text in einer bestimmten Reihenfolge angeordnet. Zwischen den Daten stehen normierte Deskriptionszeichen, z. B. steht vor der Angabe verantwortlicher Personen oder Körperschaften ein Schrägstrich, zwischen Erscheinungsort und Verlag ein Doppelpunkt. Angaben, die nicht aus der Vorlage stammen, sondern ermittelt wurden, sind in eckige Klammern gesetzt. Der Artikel am Anfang des Sachtitels wird in vielen bibliographischen Datenbanken bei der Sortierung übergangen. Bis heute bieten manche bibliographischen Verzeichnisse die Möglichkeit, Datensätze im ISBD-Format anzuzeigen. Die folgenden Beispiele stammen – wie Abb. 9 – aus dem Verbundkatalog des Südwestdeutschen Bibliotheksverbunds.

Brecht, Bertolt [1898–1956]

Der kaukasische Kreidekreis / Bertold Brecht. Mit einem Kommentar von Ana Kugli. – Orig.-Ausg., 1. Aufl., [Nachdr.]. – Frankfurt am Main: Suhrkamp, 2007. – 188 S. (Suhrkamp-BasisBibliothek ; 42 : Text und Kommentar)

ISBN 3-518-18842-9

MARC 21

Unabhängig von ihrer Anzeige auf dem Bildschirm sind die Datensätze selbst in einem maschinenlesbaren Austauschformat gespeichert. International gebräuchlich ist das von der Library of Congress entwickelte MARC-Format (Machine-Readable Cataloging, MARC 21); im deutschsprachigen Bereich wird es vielfach mit kleineren Modifikationen genutzt. Ein MARC-Datensatz besteht aus Kategorien, die mit 3-stelligen Feldnummern gekennzeichnet sind und in Unterfelder gegliedert sein können. Zur Steuerung bestimmter Funktionen bei der Katalogisierung werden Indikatoren in Form von Zahlen und Zeichen verwendet. Auch dieses Format können sich Nutzer in einigen Fällen bei der Trefferansicht anzeigen lassen. Der Vorteil ist, dass die kompletten, oft detailreichen Erschließungsdaten sichtbar und vielfach für die Nachnutzung angeboten werden.

Tab.1:

020		\$a3518188429\$93-518-18842-9
100	1	\$aBrecht, Bertolt\$d1898-1956\$0(DE-588)118514768\$0(DE-627)133840107\$0(DE-576)160414121\$4aut
245	14	\$aDer kaukasische Kreidekreis\$cBertold Brecht. Mit einem Kommentar von Ana Kugli
246	30	\$aDer vorliegende Text folgt der Ausgabe "Bertolt Brecht: Werke. Große kommentierte Berliner und Frankfurter Ausgabe. Hrsg. von Werner Hecht [u. a.] Bd. 8: Stücke 8. Bearb. von Klaus-Detlef Müller. Berlin, Weimar/Frankfurt/M., 1992, S. 93–185.
250		\$aOrig.-Ausg., 1. Aufl., [Nachdr.]
264	1	\$aFrankfurt am Main\$bSuhrkamp\$c2007
300		\$a188 S.
490	1	\$aSuhrkamp BasisBibliothek\$aText und Kommentar\$v42

Überaus wichtig für die maschinelle Übernahme von Metadaten digitaler Objekte sind auch die ebenfalls von der Library of Congress entwickelten XML-basierten Standards *Metadata Encoding and Transmission Format* (METS) und *Metadata Object Description Schema* (MODS). METS beschreibt die administrativen und strukturellen Eigenschaften von Digitalisaten, MODS umfasst die beschreibenden bibliografischen Metadaten. Das Datenmodell *Resource Description Framework* (RDF) dient vor allem dazu, die Art von Beziehungen zwischen Objekten zu verdeutlichen und bildet daher eine wichtige Grundlage für das *Semantic Web*.

Dublin Core

Für die Erschließung von Websites und anderen elektronischen Ressourcen, insbesondere auch in Digitalen Bibliotheken, wurde das Dublin Core Metadata Set konzipiert. Der Begriff entstammt der Dublin Core Metadata Initiative (DCMI), die Mitte der 1990er Jahre in Dublin, Ohio, gegründet wurde. Das Schema ist bewusst sehr einfach gehalten und sollte ursprünglich den Autoren von Web-Ressourcen dazu dienen, ihre Dokumente selbst zu katalogisieren. Die ursprüngliche Version von Dublin Core besteht aus 15 Grund-Kategorien (Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights); in abgewandelten Formen ist Dublin Core bis heute weit verbreitet. Die Reihenfolge der Kategorien ist nicht fest vorgeschrieben. Ebenso wenig ist die Form festgelegt, in der die einzelnen Daten angegeben werden. Es gibt jedoch Empfehlungen zur Verwendung adäquater Standards. Die bibliothekarischen Anwender orientieren sich an den in ihrem Bereich üblichen Regeln. Zum Teil können die Daten jedoch auch automatisch generiert werden. Neben dem vollen, im Hintergrund gespeicherten Format wird in der Trefferliste meist ein übersichtlicheres Kurzformat angezeigt, das nur die wichtigsten Informationen wie Verfasser, Titel und Erscheinungsjahr enthält.

d) Datenanzeige und Ranking


Durch die dynamischen Aufbereitungsmöglichkeiten von Daten bieten heutige Informationsressourcen auch eine Vielzahl von Möglichkeiten in Bezug auf die Darstellung der Rechercheergebnisse und deren Ranking.

Kurztitelanzeige

Da die Inhalte moderner Informationssysteme qualitativ und quantitativ immer umfangreicher werden, wäre die vollständige Ansicht aller Inhalte nach der Recherche sehr unübersichtlich und würde einen wertenden Gesamtüberblick über die Ergebnisse nahezu unmöglich machen. Daher präsentieren fast alle Ressourcen die Ergebnisse einer Suche zunächst in Form einer Kurztitelliste, in der von jedem Treffer nur diejenigen elementaren Daten dargestellt werden, die eine rasche kursorische Einordnung des Treffers ermöglichen. Bei bibliographischen Datensätzen betrifft dies häufig die Kategorien Autor, Titel (zum Teil gekürzt), Erscheinungsjahr und Medienart (vielfach als Symbol dargestellt). Häufig ebenfalls bereits an dieser Stelle dargestellt ist die Möglichkeit des Zugangs (Anzeige verschiedener Exemplare, Bestellbutton, Link zum Volltext, zur Fern-

leihe etc.) sowie die Möglichkeit, Treffer abzuspeichern oder zu exportieren. Weitere Informationen, wie einleitende Sätze eines Klappentexts oder Abstracts, können ebenfalls bereits an dieser Stelle angezeigt werden.

☐ 6



Buch

Praxishandbuch Forschungsdatenmanagement

von Neumann, Janna Putnings, Markus Neuroth, Heike

2021

„...Der Aufbau der nationalen Forschungsdateninfrastruktur und European Open Science Cloud stellt die Anbieter und Nutzer von Forschungsdaten vor technische aber...“

Volltext

zwischenspeichern

in Merkliste speichern

Abb. 10: Kurztitelanzeige aus dem Katalog der ULB Bonn

Die Unterscheidung zwischen Kurztitelanzeige und Vollanzeige findet sich nicht nur bei bibliographischen Datenbanken, sondern auch bei vielen anderen Informationssystemen, z. B. biographischen Verzeichnissen und Bilddatenbanken oder bei Suchmaschinen; wichtig ist diese Form der Darstellung vor allem immer dann, wenn eine Suche eine Vielzahl von Treffern ergibt.

Volltitelanzeige

Die Anzeige von weitaus umfassenderen Daten eines Datensatzes bietet die Volltitelanzeige. Sie kann in der Art der Darstellung sehr unterschiedlich gestaltet sein und präsentiert eine Vielzahl oder auch alle zu einem Datensatz gehörenden bzw. im System vorhandenen Bestandteile. Dies betrifft bei bibliographischen Datenbanken vor allem alle klassischen beschreibenden Metadaten, vielfach aber auch administrative und technische Daten sowie Daten, die der internen Verarbeitung des Datensatzes dienen. Da insbesondere interne Daten für den Nutzer von Informationsressourcen nicht immer von Interesse sind, werden diese – zumindest in der Standardeinstellung – vielfach auch nicht dargestellt. Oft lässt sich die Art der Anzeige der Trefferdaten sowie der Umfang der angezeigten Daten auch individuell gestalten (s. o.); dies gilt auch für den Datenexport, auch für diese Funktion können vielfach unterschiedliche technische Datenformate, Anzeigeformen und Datenmengen ausgewählt werden.

Wichtige zusätzliche Informationen, die sich heute in bibliographischen Datensätzen vielfach finden, sind Angaben aus dem Bereich der *Zitationsanalyse*, also verlinkte Angaben darüber, welche weiteren Titel von einer Publikation zitiert werden (Zitate) und von welchen diese Publikation zitiert wird (Zitationen); abgeleitet aus diesen Daten finden sich vielfach auch weitere Funktionen, wie

Verweise auf inhaltlich vergleichbare Titel und Ergebnisse bibliographischer *Metriken* (Impact Factor, h-Faktor etc., s. u. S. 292). Liegt ein Werk im Volltext vor, finden sich vielfach auch *statistische Angaben* über die verschiedenen Formen der Nutzung: Wie oft wurde ein bibliographischer Datensatz bzw. ein Volltext aufgerufen bzw. exportiert etc.? Zum Teil werden auch weitere graphisch aufbereitete Informationen zum Autor angeboten: Wann hat er, gemeinsam mit welchen anderen Autoren, welche Werke in welchen Zeitschriften zu welchen Fachgebieten veröffentlicht? Ebenfalls in der Volltitelanzeige angezeigt oder dort aufrufbar sind die verschiedenen Elemente der *Kataloganreicherung* (Abbildungen der Publikation, Inhaltsverzeichnisse, Abstracts, Rezensionen etc., s. u. S. 98). Auch die vielfältigen *Verlinkungen* in der Volltitelanzeige bieten umfassende Möglichkeit zur Anzeige weiterer Informationen oder zu weiterführenden Recherchen. Vielfach führen Links zu internen oder externen Informationen z. B. über Autoren, Schlagwörter, Verlage, Zeitschriften etc. Regelmäßig sind typische Suchbegriffe wie Autor, Schlagwort und Notation einer Klassifikation auch als Link gestaltet, so dass das Anklicken dieser Links unkompliziert eine Anschlussrecherche ermöglicht: Welche Veröffentlichungen eines Autors sind in der Datenbank verzeichnet, welche sachlich passenden Dokumente finden sich zu einem Schlagwort bzw. zu einer Notation?

Visuelle Darstellung von Treffermengen und Beziehungen

Moderne Datenbanken bieten vielfältige, interaktive Suchfunktionen und Darstellungsmöglichkeiten. So lassen sich Daten mit geographischen Bezügen vielfach sowohl über virtuelle Landkarten recherchieren, aber auch die Treffermengen lassen sich vielfach als Einträge auf Landkarten darstellen. Weitere Möglichkeiten der Visualisierung von Suchergebnissen bieten ihre Darstellung in Form von Kreis- oder Flächendiagrammen mit unterschiedlichem Informationsgehalt. So lassen sich bei der Suche nach den Publikationen eines Autors die Treffer beispielsweise als Flächen darstellen, in denen die inhaltliche Ausrichtung der Titel, die Sprache der Publikation oder die Zeitschrift, in der sie erschienen sind, farblich abgesetzt dargestellt ist. Thematische Beziehungen werden vielfach mit Hilfe von Verbindungen zwischen Begriffen, Personen oder Orten visualisiert; die Intensität oder Qualität von Beziehungen kann zusätzlich durch die Größe der Begriffe (Prinzip der Wortwolke, Tag Cloud) oder durch farbliche Unterscheidungen dargestellt werden. In einigen Fällen lassen sich auf diese Weise umfassende, interaktive Beziehungsnetzwerke darstellen, zum Teil auch in Form von dreidimensionalen Darstellungen. Solche visualisierten Darstellungsformen ermöglichen einen schnellen Überblick über verschiedene Aspekte von zum Teil

umfassenden Datenmengen; darüber hinaus erlauben sie, Rechercheanfragen intuitiv zu präzisieren oder umzuformulieren, indem ein anderer Begriff ins Zentrum des Netzwerks gerückt wird.

Ranking

Rankingsysteme, die möglichst nutzerrelevante Inhalte repräsentieren wollen, spielen nicht nur bei Suchmaschinen im Internet und in sozialen Netzwerken, sondern bei allen Typen von Informationsressourcen eine große Rolle. Für die Sortierung von Suchergebnissen bieten Informationssysteme in der Regel verschiedene Einstellungen an. Voreingestellt ist meist eine Relevanzsortierung, bei der die mutmaßlich passendsten Ergebnisse an der Spitze der Trefferliste erscheinen. Grundsätzlich gilt hierbei, dass die Qualität eines solchen Relevanzrankings mit der Menge, Vielfalt und Qualität der Daten zunimmt, die hierfür analysiert werden können. Aus diesem Grund bieten Internetsuchmaschinen in der Regel eine deutlich bessere Relevanzsortierung als Volltextdatenbanken, diese wiederum bessere Ergebnisse als Referenzdatenbanken, die lediglich die bibliographischen Daten der verzeichneten Publikationen enthalten.

Bibliographische Datenbanken berücksichtigen für die Relevanzsortierung vor allem Faktoren wie das Ketegorienfeld und die Häufigkeit des Vorkommens, den Abstand der Suchbegriffe sowie die Aktualität der Titel. Bei Discovery-Systemen (s. u. S. 94) werden zum Teil auch zusätzliche Faktoren wie Verlag und Beliebtheit des Titels eingesetzt. In vielen Fällen sind diese Kriterien nicht ausreichend für eine sinnvolle Sortierung; die Ergebnisse bleiben oft problematisch. Gerade bei diesen Ressourcen empfiehlt es sich daher dringend, sich bei einer Durchsicht langer Trefferlisten nicht nur – wie dies bei den Ergebnissen von Suchmaschinen vielfach sinnvoll ist – auf die ersten zwei bis drei Seiten zu konzentrieren, sondern die Suchergebnisse umfassender oder auch vollständig auf ihre Relevanz zu prüfen. Vielfach steht mit der Facettierung eine weitere Möglichkeit zur Verfügung, um Trefferlisten sinnvoll einzugrenzen (s. u.). In Einzelfällen kann es auch sinnvoller sein, das Ranking auf eine chronologische Anzeige der Datensätze umzuschalten. Eine alphabetische Sortierung empfiehlt sich vor allem, wenn Titellisten erstellt werden sollen. Bei der alphabetischen Sortierung nach Titeln werden üblicherweise zuerst alle Titel angezeigt, die mit einem Zeichen oder einer Zahl beginnen.

Die komplexen Methoden der Relevanzsortierung kommerzieller Suchmaschinen werden von den Betreibern in der Regel nicht offengelegt, sie bilden ein wichtiges Element für den Erfolg der Suchmaschine. Zudem verringert ihre Ge-

heimhaltung, dass die Anbieter von Webseiten durch Suchmaschinenoptimierung (Search Engine Optimization, SEO) ihrer Seiten die Positionierung in der Trefferanzeige manipulieren. So berücksichtigt z. B. *Google* eine Vielzahl von Faktoren – dazu gehören die Analyse der Texte, die Anzahl der Links auf diese Seite und ihrer Verlinkungen, die Sprache, die Domain, die Struktur, die bisherige Nutzung und die Benutzungsfreundlichkeit (User Experience). Die Faktoren werden unterschiedlich gewichtet, kombiniert und insgesamt beständig weiterentwickelt.

Filter und Facetten

Eine weitere Möglichkeit, Suchergebnisse einzugrenzen bzw. zu präzisieren, bieten Filter und Facetten. *Filter* werden dabei in der Regel bereits vor dem Absenden der Suchanfrage aktiviert; sie beziehen sich zumeist auf die Kriterien Erscheinungszeitraum, Publikationsart oder auch Sachgruppen. Von *Facetten* bzw. der Facettierung von Suchergebnissen spricht man in der Regel bei der nachträglichen Eingrenzung einer Trefferliste. Hierfür werden die Ergebnisse einer Trefferliste automatisch analysiert und in der Trefferanzeige wird dem Nutzer dann die Möglichkeit gegeben, Trefferlisten durch die Auswahl von Facetten kontextabhängig weiter einzugrenzen. Häufig angebotene Facetten sind z. B. Zugriffsmöglichkeit, Schlagwort, Dokumenttyp, Erscheinungsjahr, Sprache, Autor oder Institution. Bei den angebotenen Facetten wird bereits die Anzahl der Dokumente angezeigt, was z. B. Rückschlüsse darüber erlaubt, in welchem Erscheinungszeitraum, in welcher Sprache, von welchen Personen oder Institutionen zu dem Thema am meisten publiziert wurde. Ein schrittweises Eingrenzen der Treffermenge wird als *drill down* bezeichnet. Tatsächlich ist jedoch festzustellen, dass eine strenge Unterscheidung von Filter und Facetten vielfach nicht vorgenommen wird; häufig werden mit den Begriffen Filter und filtern auch Facetten und facetten bezeichnet. Manche Informationssysteme bieten ihren Nutzern auch die Funktion, innerhalb einer bestehenden Trefferliste mit neuen Suchbegriffen weiterzusuchen.

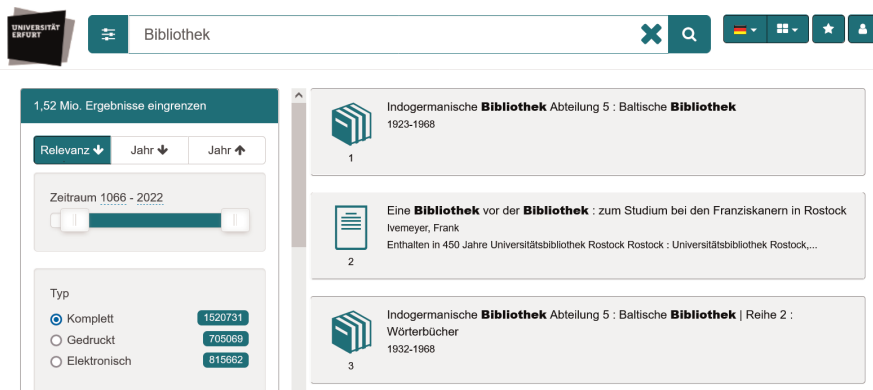


Abb. 11: Kurztitelanzeige im Discovery-System der UB Erfurt mit Einstellungsmöglichkeit des Rankings und der Facettierungsfunktion

VI Datenexport, Datenverwaltung und Dokumentenzugang

Mit einer erfolgreichen Suche in einer einschlägigen Informationsressource ist ein Recherchevorgang noch nicht abgeschlossen. Regelmäßig geht es im Anschluss darum, interessante Informationen, oft bibliographische Datensätze, aus einer Trefferliste auszuwählen, sie abzuspeichern und für die folgende Verarbeitung und Verwaltung zu exportieren. Eine weitere Aufgabe besteht darin, ausgehend von bibliographischen Treffern Zugriff auf die zugehörigen Volltexte zu erlangen. Während dies bei Bibliothekskatalogen überaus unkompliziert ist – die Titel können unmittelbar am Regal entnommen, bestellt bzw. vorbestellt oder online aufgerufen werden –, stellt gerade dieser Schritt bei den vielen nicht bestandsbezogenen Ressourcen (Bibliographien, Fachportalen, Suchmaschinen etc.) kein geringes Problem dar.

a) Datenspeicherung, Datenexport und Literaturverwaltung

Sehr viele Informationsressourcen bieten heute die Möglichkeit, Daten zu exportieren und individuell in anderen Kontexten weiter zu nutzen und zu verarbeiten. Fast immer trifft dies auf Metadaten zu, vielfach jedoch auch auf angebote-