

Stand: 7.2.2024

# Skript Datenbankwissen

## Warum erste Theorie?

Es gibt weltweit tausende Datenbanken, und eigentlich wäre es theoretisch gut, wenn man jeweils eine spezifische Einführung in eine einzelne Datenbank erhalten würde. In der Praxis wird es jedoch darauf hinauslaufen, dass man höchstens eine Einführung in die Datenbank besucht, die für das eigene Fach am wichtigsten ist. Das ist aber nicht weiter schlimm, weil Datenbanken alle den gleichen Prinzipien folgen. Wenn man diese Prinzipien einmal kennt, kann man sich auch leicht in Datenbanken einarbeiten, mit denen man noch nicht gearbeitet hat. Das ist vergleichbar mit dem Autofahren. Wenn man weiß, dass alle Autos ein Lenkrad, ein Gaspedal, eine Bremse und ein wie auch immer gestaltetes Schaltgetriebe haben, kann man auch Fahrzeuge verschiedener Hersteller bedienen. Das ist bei Datenbanken nicht anders. Bei Datenbanken hat man zudem den Vorteil, dass es auf dem Markt zwei große Anbieter gibt, die Datenbanken hosten: Ebsco und ProQuest. Sehr häufig werden einem daher entweder die Oberflächen von Ebsco oder ProQuest begegnen, wenn man eine Datenbank nutzt. Das führt dazu, dass man sich nicht immer komplett neu in eine Oberfläche einarbeiten muss. Ganz statisch bleiben jedoch auch die Oberflächen von Ebsco und ProQuest nicht, da beide ihr Angebot immer weiter entwickeln. Im Laufe der Zeit verändern sich also auch deren Oberflächen leicht, wenn neue Suchmöglichkeiten oder Features dazu kommen. Davon sollte man sich nicht verwirren lassen, sondern die Veränderungen eher als Aufforderung ansehen, etwas nützliches Neues zu entdecken.

## Unterschied Bibliothekskatalog - Datenbank

Bevor man weiter auf Datenbanken eingeht, sollte man den Unterschied zwischen einem Bibliothekskatalog und einer Datenbank aufzeigen. Die beiden unterscheiden sich besonders in zwei Dingen: der Verfügbarkeit (wo man etwas findet) und dem Inhalt (was man jeweils darin findet).

Eine Datenbank ist ein Verzeichnis von spezieller Literatur ohne Rücksicht auf Vorhandensein in einer bestimmten Bibliothek (also alles, was es weltweit gibt, ohne dass die Literatur auch in Köln vorhanden sein muss). Ein Bibliothekskatalog ist das Bestandsverzeichnis einer bestimmten Bibliothek (also nur alles, was es an der Uni Köln gibt).

In einer Datenbank findet man Monographien (also Bücher), Zeitschriftenaufsätze und andere Arten von Literatur. In einem Bibliothekskatalog findet man vor allem Monographien. In der Regel sind keine einzelnen Aufsätze suchbar (es kann sein, dass vereinzelt Aufsätze in Bibliothekskatalogen auftauchen, weil diese Regel nicht von allen Bibliotheken strikt eingehalten wird. Die Norm ist dies aber nicht und eine Suche in Datenbanken wird damit auch nicht ersetzt).

Daher kann man mit Autor- und Titelangaben im OPAC regelmäßig nur nach Büchern suchen, aber nicht nach Zeitschriften-Aufsätzen.

### **Unterschied Stichwort - Schlagwort**

Die Datenbankanbieter haben Redaktionen, die alle relevanten Quellen (Zeitschriften, Dissertationen, Monographien etc.) auswerten und erschließen (daher sind viele Datenbanken auch so teuer). Mit Erschließen ist gemeint, dass jemand den Inhalt dieser Zeitschriftenaufsätze, Dissertationen oder Monographien beschreibt und sich dabei eines normierten (also nicht beliebigen) Vokabulars bedient. Anders ausgedrückt, jemand vergibt Schlagwörter für diese Literaturquellen. Ein Schlagwort ist ein normierter Begriff zur Bezeichnung von Dokumenteninhalten. Das kann man sich in etwa wie eine Schublade vorstellen, in die Fundstücke zum gleichen Thema gelegt werden. Die Beschriftung dieser Schublade ist dann das Schlagwort. Wenn ich also, um im Bild zu bleiben, die Schublade öffne, finde ich dort alle Literatur zu dem Thema und zwar unabhängig von der Literaturform (also Zeitschriftenaufsätze, Monographien, Dissertationen und andere Literaturformen) und auch unabhängig von der Sprache, in der diese Literaturquellen verfasst wurden (also Literatur auf Deutsch, Englisch, Französisch etc.).

Das Gegenteil des Schlagworts ist das Stichwort. Hier hat sich kein Mensch Gedanken zum Dokumenteninhalt gemacht. Hier sucht man dann einfach nach einem Stichwort, das hoffentlich irgendwo in dem Datensatz benutzt wird. Das kann man sich so vorstellen wie eine Suche in einem großen Word-Dokument. Da gibt man ein Stichwort ein und bekommt alle Treffer angezeigt, wo sich dieses Stichwort im Text finden lässt. Das bedeutet aber nicht, dass ich alle Stellen im Text finde, die mich interessieren. Ich kann auch wichtige Stellen im Text verpassen, einfach, weil dieses Stichwort dort nicht benutzt wird (vielleicht weil ein alternativer Begriff verwendet wird, zum Beispiel „Kraftfahrzeug“ statt „Auto“).

Um den Unterschied zwischen Stichwort und Schlagwort in der Literatursuche einmal mit einem Beispiel zu beleuchten: Eines der Standardwerke über die USA zur Zeit des Amerikanischen Bürgerkriegs ist „The Battle Cry of Freedom: the Civil War Era“ von James M. McPherson. Wenn man nun einfach mit dem Stichwort „Amerikanischer Bürgerkrieg“ sucht, findet man das Buch nicht, denn im Titel kommt eben nicht der deutsche Begriff „Amerikanischer Bürgerkrieg“ vor. Aber auch mit dem englischsprachigen Stichwort „American Civil War“ würde man das Buch nicht unbedingt finden.

In den Datenbanken, die meist englisch-sprachig sind, wird für Schlagwort meist der Begriff „subject“ (in manchen auch "Thesaurus Term" oder „Descriptor“) und für Stichwort „keyword“ benutzt. Je nach Datenbank kann man die Stichwortsuche auf den Volltext der Aufsätze ausdehnen. Datenbanken enthalten oft Abstracts zu den Literaturquellen, wodurch eine inhaltliche Suche mit einem Stichwort erfolgversprechender wird.

Schlagwörter werden in Datendanken häufig auch in verschiedene Kategorien eingeteilt: Sachschlagwörter (z.B. „detective fiction“), Geographische Schlagwörter (z.B. „Berlin“) oder

zeitliche Schlagwörter (z.B. „19th century“). Jede dieser Kategorien hat dann ihren eigenen Thesaurus/Index.

### **Schlagwortindex /Thesaurus**

Jede Datenbank hat eine Liste mit den Schlagwörtern, die in dieser Datenbank benutzt werden. Diese Liste nennt sich Thesaurus. Da die großen Datenbankanbieter im englischsprachigen Raum beheimatet sind, ist die Datenbanksprache in der Regel Englisch. Das bedeutet, dass auch nur mit englischsprachigen Schlagwörtern etwas gefunden werden kann. Also findet man im Thesaurus nichts mit „Kriminalroman“, sondern eher mit „detective novel“ oder „crime novel“. Dass man hier zwei mögliche englischsprachige Schlagwörter angibt, hat seine Richtigkeit und soll auch Folgendes verdeutlichen: Manche Begriffe sind nicht genau festgelegt, daher sollte man immer in den Thesaurus schauen, welchen Begriff die Datenbank kennt. Eine Datenbank ist zunächst eine Software und die macht nur das, was man ihr sagt. Wenn man einfach nach „crime novel“ sucht und die Datenbank verwendet „detective fiction“, dann bekommt man einfach keine Treffer. Die Datenbank wird keinen Hinweis geben, dass man mit dem anderen Begriff etwas gefunden hätte. Daher ist es immer sinnvoll, sich vor einer Suche den Thesaurus anzuschauen. Ein Blick in den Thesaurus ist auch deswegen sinnvoll, weil bei den Schlagwörtern häufig Ober- und Unterbegriffe genannt werden, die dieses Schlagwort in einen Kontext stellen und so gerade beim Verstehen von Themen, die neu für einen sind, helfen. Im Screenshot sieht man hinter den Schlagwörtern kleine Kästchen, die man anklicken kann, um Ober- und Unterbegriffe angezeigt zu bekommen. Häufig werden dort auch verwandte Begriffe angezeigt. Um im Beispiel von oben zu bleiben, könnte man einen Verweis finden, dass man statt „crime novel“ das verwandte Schlagwort „detective fiction“ verwenden sollte.

### **Boolesche Operatoren**

Boolesche Operatoren kennt fast niemand, auch wenn fast jeder mindestens einen dieser Verknüpfungsoperatoren ständig benutzt. Bei der Suche in einem Bibliothekskatalog mit Autor und Titel wird automatisch eine „UND“- Verknüpfung gesetzt (damit man nicht alle Bücher eines Autors angezeigt bekommt oder alle Bücher mit gleichem Titel von verschiedenen Autoren). Auch in Datenbanken wird überwiegend der Operator „UND“ benutzt (weil man mit der Verknüpfung mehrerer Schlagwörter eine relevantere Treffermenge erhält). Bei dem Beispiel auf Folie 9 soll die Suche Treffer zum Thema „Shakespeares Komödien“ ergeben. Wenn man nur das Schlagwort Shakespeare benutzte, würde es eine sehr große Anzahl an Treffern geben, genauso wie bei der ausschließlichen Verwendung des Schlagworts „Comedy“. Mit der „UND“-Verknüpfung findet man gezielt die Schnittmenge der beiden Themen und somit weniger, dafür aber relevantere Treffer. Neben dem am häufigsten genutzten Operator „UND“ gibt es auch die beiden viel weniger häufig benutzten Operatoren „ODER“ und „NICHT“

Bei dem Beispiel auf Folie 10 wird der Operator „ODER“ benutzt. Hierbei kann man bei einer einzigen Suchabfrage zwei Suchbegriffe abfragen. Das bietet sich besonders für ähnliche und

verwandte Begriffe an. Das spart Zeit, weil die Treffer in einer Liste ausgegeben werden und auch vorab möglicherweise doppelt vorhandene Treffer (sogenannte Dubletten) in dieser Liste bereinigt werden.

Bei dem Beispiel auf Folie 11 wird der Operator „NICHT“ benutzt. Dies wird auch eher selten verwendet und dann dafür, ein großes Thema einzugrenzen. Im vorliegenden Fall soll das große Thema „Shakespeare“ um den Aspekt „Komödie“ eingegrenzt werden. Man möchte also alles finden, was sich zwar mit Shakespeare, aber nicht mit seinen Komödien beschäftigt.

### **Trunkieren**

Bei einer Suche kann man auch einen oder mehrere Buchstaben durch Symbole wie \*, ? oder \$ ersetzen, und zwar sowohl vor als auch nach dem Wortstamm. Damit kann man mit einer einzelnen Suche viele Kombinationen abfragen. So findet etwa „\*sprachigkeit“ „Mehrsprachigkeit“ und „Zweisprachigkeit“ oder „comput\*“ findet „computability“, „computer“ und „computing“. Dieses Trunkieren genannte Verfahren hilft ebenfalls dabei, bei der Suche Zeit zu sparen. Je nach Datenbank wird ein anderes Trunkierungssymbol verwendet. Hier muss man einfach auf die Hilfeseiten der Datenbank schauen und sich dort informieren, welches Symbol in der vorliegenden Datenbank benutzt wird. Es gibt noch eine Einschränkung: die Trunkierung vor dem Wortstamm funktioniert nicht in jeder Datenbank.

Das Beispiel Trunkieren macht auch anschaulich, dass man mit theoretischem Wissen über Datenbanken mehr aus diesen herausholen kann. Wenn man nicht von der Trunkierungsfunktion wüsste, könnte man diese nicht nutzen und damit auch nicht so viel aus der Datenbank entlocken, wie es eigentlich möglich ist.

### **Auswerten der Treffermenge**

Datenbanken bieten eine Vielzahl von Möglichkeiten, die gefundene Treffermenge auszuwerten und einzugrenzen. Hier bieten sich sachliche wie inhaltliche Kriterien an. Sachliche Kriterien können das Publikationsjahr (z.B. nur die letzten fünf Jahre), der Dokumententyp (z.B. nur Zeitschriftenartikel, keine Monographien) oder die Sprache sein, in der etwas veröffentlicht wurde (z.B. nur Artikel auf Deutsch und Englisch). Inhaltliche Kriterien können etwa die Schlagwörter sein, mit denen die Treffer verknüpft sind. Eine Literaturquelle kann mit vielen Schlagwörtern verknüpft sein, nicht nur mit denen, nach denen gesucht wurde. Die Datenbanken von Ebsco und ProQuest werten daher die Treffermenge dahingehend aus, dass eine Liste mit den am häufigsten verwendeten Schlagwörtern dieser Treffermenge angegeben wird. Diese einzelnen Schlagwörter können nun genutzt werden, die Treffermenge einzugrenzen, indem sie zum Filterkriterium gemacht werden (entweder als Positivkriterium („include“) oder als Negativkriterium („exclude“). Eine weitere inhaltliche Eingrenzung wäre das Kriterium „peer reviewed“, also nur solche Treffer anzuzeigen, die eine Qualitätssicherung durch Begutachtung durch Fachkolleg\*innen aufweisen können. Allen Eingrenzungswerkzeugen ist gemeinsam, dass sie dazu dienen

sollen, die Treffermenge noch passgenauer zu beschneiden und so besser verwertbar zu machen.

### **Schneeballsystem: References / cited by**

Bei den einzelnen Treffern in den Datenbanken wird häufig (besonders bei neueren Veröffentlichungen) angegeben, welche Literatur für diese Publikation benutzt wurde. Das wird dort unter „References“ angegeben. Anders ausgedrückt, muss man gar nicht erst in den Aufsatz oder das Buch schauen, um das Literaturverzeichnis einzusehen. Denn das Literaturverzeichnis ist in dem Datensatz gleich angegeben (meist als Hyperlink. Wenn man dann darauf klickt, kommt man direkt zu diesem Datensatz in der Datenbank.).

Diese Vorgehensweise (ich nehme mir einen möglichst aktuellen Aufsatz oder ein Buch und schaue im Literaturverzeichnis, was der Autor/die Autorin als Sekundärliteratur benutzt hat) gab es auch schon zu den alten vor-digitalen Zeiten. Der Name für dieses Verfahren ist „Schneeballsystem“. Ähnlich wie ein Schneeball, den man einen Hügel hinabrollen lässt, vergrößert sich auch das Suchergebnis, wenn man dieses Verfahren immer wieder anwendet (aus dem Inhaltverzeichnis der neuesten Publikation nimmt man z.B. drei ältere Publikationen, aus diesen drei dann z.B. wieder jeweils drei ältere Publikationen usw.). Auf diese Weise erhält man recht schnell einen Überblick über die für einen relevante Literatur. Man sollte jedoch im Hinterkopf behalten, dass man sich diese Schnelligkeit mit ein paar Nachteilen erkauft. So ist die Suche nicht so umfassend wie eine Datenbankrecherche und man erhält nur Literatur, die vor einem bestimmten Datum (eben dem der jüngsten Literaturquelle, die man gefunden hat) veröffentlicht wurden. Zudem besteht auch die Gefahr, dass man sich nur innerhalb einer bestimmten Denkschule bewegt und andere Denkrichtungen nicht wahrnimmt (wenn der Autor einer bestimmten Denkschule angehört, wird er oder sie unter Umständen auch nur solche Quellen zitieren, die zu dieser Denkschule passen – wenn man also gar nicht mit dem Thema vertraut ist, kann man leicht in diese Fall tappen.)

Eine Art Schneeballsystem in die Zukunft bieten Datenbanken auch an. In der „Cited by“ (oder so ähnlich) genannten Funktion wird angegeben, wer später diesen Beitrag zitiert hat. Wenn ich also einen Aufsatz oder ein Buch in der Datenbank gefunden habe, sehe ich, welche Aufsätze oder Bücher wiederum diese Quelle benutzt haben. Wenn eine Literaturquelle häufig zitiert wird, spricht das für die Qualität dieses Aufsatzes oder Buches. Auch bei diesem Schneeballsystem in die Zukunft gelten die oben genannten Einschränkungen: die Suche ist nicht so umfassend wie eine Datenbankrecherche und es besteht die Gefahr, nur eine Denkschule zu rezipieren.

Es gibt auch eine gesonderte Form der Suche namens „Cited Reference Search“, die sich explizit damit beschäftigt, wer später einen bestimmten Beitrag zitiert hat. Wie diese „Cited Reference Search“ benutzt wird, wird in einem Video erklärt, das auch auf Ilias zu finden ist.

### **Immer auf dem Laufenden bleiben: Search Alert**

Wenn man sich bei einer Datenbank einen Account anlegt, kann man auch Suchanfragen speichern. Auf diese Weise muss man die Suche nicht immer neu formulieren, wenn man die Datenbank später wieder benutzt. Man kann diese gespeicherten Suchen auch zu einem sogenannten „Alert“ ausbauen. Hier bekommt man in selbst definierten Abständen eine Nachricht von der Datenbank, wenn neue Literaturquellen in der Datenbank eingestellt wurden, die dem Suchprofil entsprechen. Dieser Alerting-Service lohnt sich aber nur, wenn man über einen längeren Zeitraum (etwa bei einer Abschlussarbeit) die brandaktuelle Literatur zu einem Thema benötigt.

### **Literaturhinweise**

Die Literaturrecherche wird meist nur als eines von vielen Themen in den Ratgebern zum wissenschaftlichen Arbeiten behandelt. Es gibt jedoch auch Ratgeber, die sich explizit mit der Literaturrecherche beschäftigen.

Bei DeGruyter ist die Reihe „Erfolgreich Recherchieren“ erschienen:

<https://www.degruyter.com/serial/er-b/html#volumes>

Hier findet man Ratgeber für folgende Fächer: Altertumswissenschaften und Archäologie, Biowissenschaften, Wirtschaftswissenschaften, Jura, Linguistik, Informatik, Medizin, Mathematik, Kunstgeschichte, Erziehungswissenschaften, Anglistik und Amerikanistik, Politik- und Sozialwissenschaften, Romanistik, Geschichte und Germanistik.

Auch wenn die Titel zwischen 2012 und 2016 erschienen sind und daher nicht mehr ganz aktuell sind, geben sie dennoch eine vertiefte Einführung in die Literatursuche für das jeweilige Fach.

### **Übersicht über die Datenbanken an der Universität zu Köln**

Auf dieser Seite findet man eine Übersicht über alle an der Universität zu Köln verfügbaren Datenbanken: <https://www.ub.uni-koeln.de/usbportal?service=dbbrowse>

Die Datenbanken sind nach Fächern geordnet. Die für ein Fach jeweils wichtigsten Datenbanken stehen ganz oben unter der Überschrift „TOP-Datenbanken“. Wenn man eine Datenbank anklickt, öffnet sich eine eigene Seite. Dort findet man Informationen zum Inhalt und Umfang der jeweiligen Datenbank.