

# „Information Retrieval“ / Informationsabfrage – über Suchmaschinen oder Datenbanken?

## Information retrieval – search engines versus databases

Carolin Werner<sup>1</sup>, Manuela Franke<sup>2</sup>

### Abstract

In recent decades the amount of available information has increased enormously. Especially by using the internet there are innumerable possibilities of finding this mostly decentralized information. It is therefore all the more important to identify relevant and scientifically valuable information as well as to analyze one's own specific need for information. Important tools for this kind of retrieval are databases that have been quality checked and reviewed by scientific experts as well as analyzed and indexed by information specialists who evaluate the authenticity and integrity of the database's content. This article provides more details about information retrieval and discusses some of the main reasons why it is necessary to consider this important topic systematically.

### Zusammenfassung

In zunehmendem Maße stieg in den letzten Jahrzehnten die Menge an verfügbaren Informationen und wird auch noch weiter ansteigen. Vor allem durch das Internet liegen diese in einer kaum zu erfassenden Anzahl häufig dezentralisiert vor. Umso wichtiger wird es, die für den eigenen Informationsbedarf erforderlichen Informationen zu filtern und auf ihren wissenschaftlichen Wert, die Exaktheit, hin zu prüfen. Ein wichtiges Instrument dabei sind von wissenschaftlichen Experten qualitätsgeprüfte Datenbanken, deren Dokumentinhalte von ausgebildeten Informationsspezialisten sowohl formell als auch inhaltlich erschlossen werden, um eine gezielte Recherche zu ermöglichen. Im vorliegenden Artikel soll diese Möglichkeit der Recherche näher erläutert und hauptsächlich Gründe aufgezeigt werden, warum es wichtig ist, sich mit der Thematik des ‚Information Retrieval‘, dem Wiederfinden beziehungsweise Abrufen von Informationen, außerhalb von Suchmaschinen zu beschäftigen.

## Einleitung

In allen Bereichen der naturwissenschaftlichen und medizinischen Forschung sind valide und aktuelle Informationen unabdingbar. Nicht selten sieht man sich einer Vielzahl von Informationsangeboten gegenüber, deren wissenschaftlicher Wert oft fraglich ist oder nicht beurteilt werden kann. Verbunden mit der rasanten Entwicklung des Internets findet man heute über Suchmaschinen, wie beispielsweise Google oder Yahoo, Informationen zu einer Vielzahl von Themen. Doch sind diese oft nicht von Experten auf ihren Inhalt geprüft oder zitierfähig. Beispiele dafür sind im Netz frei verfügbare Wissensportale, welche durch die allgemeine und auch nichtwissenschaftliche Community bearbeitet

werden können. Für Privatpersonen, die sich einen Informationsüberblick verschaffen wollen, können diese Angebote von Suchmaschinen eine wichtige Informationsquelle sein. Schwierig wird es jedoch, wenn man für wissenschaftliche Zwecke exakte und damit zuverlässige Informationen sucht. Das ist nur möglich über Beiträge, Berichte oder Ähnliches aus Datenbanken, die von Experten aus dem jeweiligen Wissenschaftsgebiet auf ihre Qualität überprüft worden sind. Generell sind vier Kriterien bei der Informationsbeschaffung essentiell: die Vollständigkeit der Informationen, die Genauigkeit der Erfassung, die Validität sowie die Aktualität, vor allem bei neuen wissenschaftlichen Erkenntnissen.

<sup>1</sup> Studentin der Informationswissenschaften, Studiengang Information und Dokumentation an der FH Potsdam, zurzeit Praktikantin im Zentrum für Krebsregisterdaten (ZfKD).

<sup>2</sup> Diplom-Dokumentarin, Medizinische Dokumentarin im Zentrum für Krebsregisterdaten (ZfKD) des RKI.

Wichtige Quellen für wissenschaftliche Informationen sind Fachzeitschriften, die ein sogenanntes ‚Peer-Review‘-Verfahren durchlaufen, das heißt, die darin publizierten Artikel werden vor Veröffentlichung auf ihre Qualität überprüft. Informationen zum Themenbereich Umweltmedizin sind in unterschiedlichsten Quellen zu finden. So ist die Umweltmedizin beispielsweise eng mit der Humanmedizin, Umwelttechnologie aber auch mit den Ernährungswissenschaften, der Biologie und der Chemie verknüpft. Häufig können jedoch keine klaren Grenzen zwischen den einzelnen Wissenschaften und deren Informationsangeboten gezogen werden. Dieser Aspekt der Interdisziplinarität ist bei einer auf Vollständigkeit angelegten Recherche sowie bei der Auswahl geeigneter Quellen immer zu beachten. Mittlerweile wird dem schon über eine enorme Auswahl an Datenbankanbietern („Hosts“) Rechnung getragen, die interdisziplinär erschlossene Datenbanken anbieten (Poetzsch 2005, S. 28). Für umweltbezogene Fragestellungen können beispielsweise die Informationsangebote des Umweltbundesamtes über <http://www.umweltbundesamt.de/service/weitere.htm> (Abrufdatum: 08.01.2013) genutzt werden.

## Datenbankauswahl

Datenbanken können eine enorme Anzahl an Dokumenten umfassen, so beinhaltet beispielsweise die Datenbank MEDLINE mehr als 19 Millionen Datensätze (Fact Sheet Medline 2011). Von Datenbankanbietern wird meistens eine große Anzahl verschiedener Datenbanken angeboten, die zum Beispiel über Metadatenbanken nach Themenbereichen oder Ähnlichem gefiltert werden können. Zudem ist es auch möglich, mehrere Datenbanken gleichzeitig zu durchsuchen, was wiederum fundierte Kenntnisse im Umgang mit Datenbanken, deren Beschreibungen oder Klassifikationen sowie Abfragesprachen („Retrieval“-Sprachen, siehe auch weiter unten) erfordert, da teilweise unterschiedliche Indices oder spezielle Suchfelder verwendet werden. Für die Auswahl einer Datenbank sollten folgende Kriterien unbedingt bedacht werden: die inhaltliche, zeitliche, sprachliche und geographische Abdeckung der Informationen in dieser Datenbank für die gewünschte Informationsabfrage, die Art der Datenbank (Faktendatenbank, Volltextdatenbank, bibliographische Datenbank, „Referral“-Datenbank), die Update-Frequenz, in der die Daten aktualisiert werden, der Umfang des Datenbestands und selbstverständlich die Kosten.

## Recherchekosten

Wissenschaftlich exakte und zuverlässige Informationen sind häufig nur über kostenpflichtige Datenbankangebote erhältlich. Die Kosten werden dabei nach verschiedenen Modellen berechnet. In vielen Datenbanken ist die eigentliche Durchführung der Recherche kostenlos, für die Anzeige und Ausgabe wird zum Teil jedoch ein Dokumentenpreis erhoben. Aufgrund möglicher anfallender Kosten muss eine Recherche effizient durchgeführt werden, was Kenntnisse der zur Verfügung stehenden Recherchemöglichkeiten beziehungsweise der „Retrieval“-Sprache erfordert. „Retrieval“-Sprachen sind aus verschiedenen Kommandos und logischen Operatoren bestehende Eingabesequenzen innerhalb eines Suchfeldes. Diese können in Abhängigkeit von dem Datenbankanbieter sehr unterschiedlich sein. Für einen Laien können daher nicht nur erhebliche Kosten bei einer Recherche entstehen, da dieser möglicherweise die verschiedenen Recherchemöglichkeiten und -tools nicht ausreichend kennt, sondern die Suche kann auch zu einem nicht zufriedenstellenden Ergebnis führen. In einigen Fällen macht es Sinn, für die Informationssuche einen Spezialisten („Information Broker“) zu engagieren, zum Beispiel, wenn der Informationsbedarf sehr speziell ist und Informationen dazu nicht einfach über Suchmaschinen oder frei zugängliche Quellen erhältlich sind. Insbesondere Patentrecherchen sollten nie ohne professionelle Unterstützung durchgeführt werden, da sie bei falschen oder fehlenden Informationen zu erheblichen wirtschaftlichen Problemen führen können.

## Rechercheablauf

Nicht zwangsläufig muss man eine Recherche in Auftrag geben, sondern kann diese auch selbst durchführen. Wichtig ist, dass man sich mit Datenbankrecherchen auskennt und seinen Informationsbedarf gezielt eingrenzt. Datenbankanbieter stellen den Nutzern auf ihren Websites genaue Beschreibungen ihrer Datenbanken sowie entsprechende Hilfen und Trainings zur Verfügung, um sich optimal auf eine Recherche vorbereiten zu können.

Vor einer Recherche sollte man die genaue wissenschaftliche Fragestellung eingrenzen. Dazu gehört die Analyse des tatsächlichen Informationsbedarfs. Eine konkrete Suchanfrage muss erstellt werden,

denn wird diese zu allgemein formuliert, erhält man zu viele irrelevante Dokumente.

**Folgende Fragen sind seitens des Informationssuchenden vorab zu klären (Poetzsch 2006):**

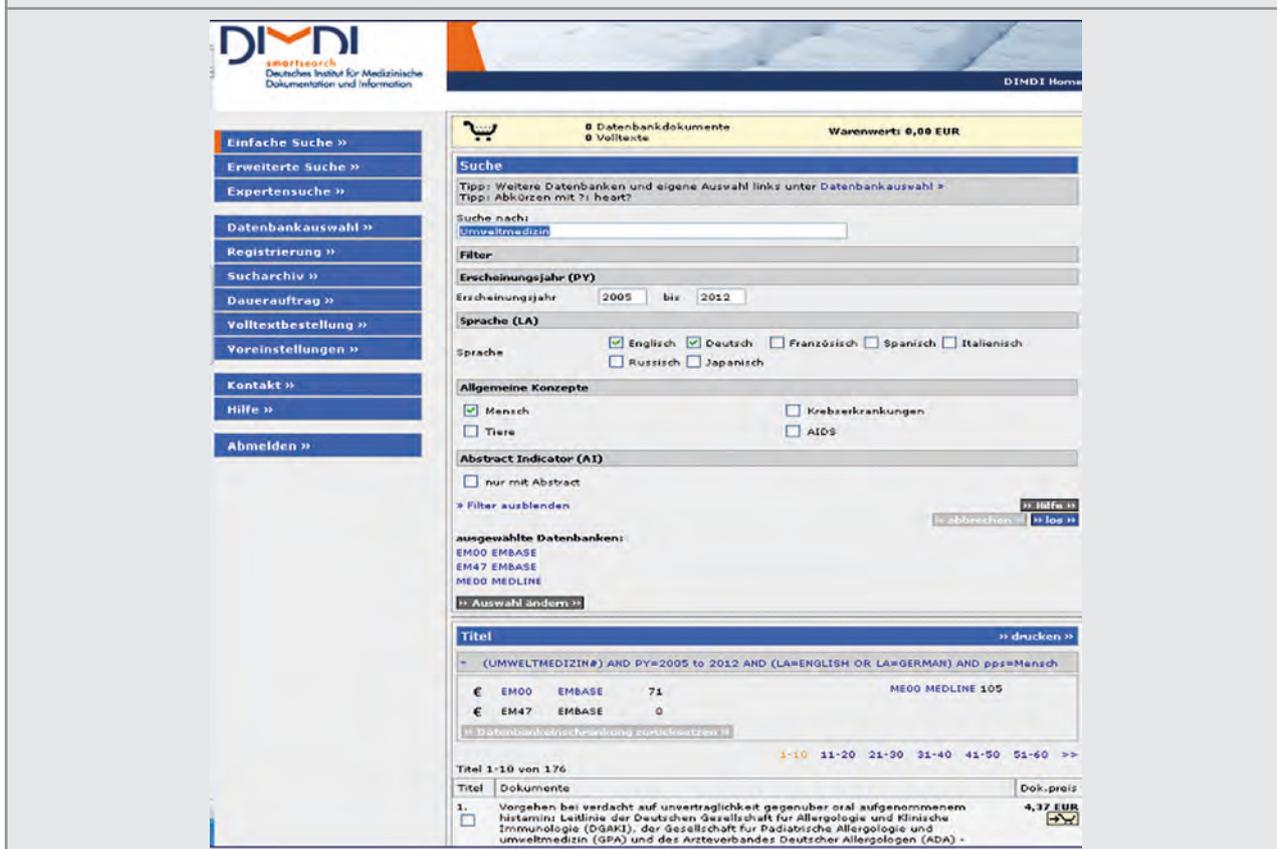
- Wie kann der Informationsbedarf eingegrenzt werden, beispielsweise über Einschränkung der Themengebiete und unter Nutzung wichtiger Schlagworte?
- Soll mit den Ergebnissen der Recherche das Thema möglichst genau bearbeitet beziehungsweise beschrieben werden oder geht es um eine möglichst vollständige Sammlung aller Dokumente, die es zu diesem Thema gibt?
- In welchem Zeitraum sollen Informationen gesucht werden, zum Beispiel in den letzten fünf Jahren?
- In welchen Sprachen können die Dokumente verfasst sein?
- In welcher Form und in welchem Format werden die Recherche-Ergebnisse benötigt (Word, PDF, Excel)?

- Werden komplette Volltexte erwartet oder reichen Abstracts beziehungsweise Literaturverweise aus?
- Bis wann werden die Recherche-Ergebnisse benötigt?
- Sind turnusmäßig über einen bestimmten Zeitraum zum gleichen Thema weitere Informationssuchen erforderlich (gegebenenfalls durch Nutzung eines Alert-Service) oder reicht eine einmalige retrospektive Recherche aus?
- Wie hoch dürfen eventuell auftretende Kosten bei der Recherche sein? Welches Budget steht zur Verfügung?

**Datenbankanbieter DIMDI**

Im Folgenden soll das Deutsche Institut für Medizinische Dokumentation und Information (DIMDI), ein Institut im Geschäftsbereich des Bundesministeriums für Gesundheit, und dessen Datenbankangebote für Recherchen als ein Beispiel zur eigenständigen Informationssuche vorgestellt wer-

Abbildung 1: Recherchemaske bei DIMDI ohne Vertrag (DIMDI 2013).  
 Quelle: [https://portal.dimdi.de/websearch/servlet/FlowController/DisplaySearchForm;jsessionid=E34B37C65109F3380A1A0D5793DE6EA2.joey#\\_\\_DEFANCHOR\\_\\_](https://portal.dimdi.de/websearch/servlet/FlowController/DisplaySearchForm;jsessionid=E34B37C65109F3380A1A0D5793DE6EA2.joey#__DEFANCHOR__) (Abrufdatum: 08.01.2013).



den. Neben diesem gibt es noch eine Vielzahl anderer Datenbankanbieter, so unter anderem Ovid ([www.ovid.com](http://www.ovid.com)), ein weiterer führender Anbieter von Informationen im Bereich Medizin. Alle Datenbankanbieter im Einzelnen zu nennen beziehungsweise vorzustellen, würde jedoch den Rahmen dieses Übersichtsbeitrages sprengen.

Das DIMDI arbeitet in seinen Aufgabenbereichen eng mit nationalen und internationalen Institutionen zusammen, unter anderem mit der Weltgesundheitsorganisation und EU-Behörden. Es betreibt ein eigenes Rechenzentrum und entwickelt moderne Software-Anwendungen für den Betrieb und den Zugriff auf die Informationssysteme und Datenbanken (siehe unter DATENBANKANBIETER). Dieses Institut bietet circa 60 Datenbanken im gesamten medizinischen Bereich und angrenzenden Gebieten für kostenfreie Recherchen und in rund 25 Datenbanken auch mit kostenfreien Datenbankdokumenten an. Für die Umweltmedizin relevante Datenbanken sind unter anderem: MEDLINE, produziert von der U.S. National Library of Medicine, EMBASE, produziert von Elsevier B.V., BIOSIS, produziert von Thomson Reuters oder das auch als Printmedium erhältliche „Deutsche Ärzteblatt“. Ohne große Vorkenntnisse kann man über die menügeführte Suchoberfläche „DIMDI SmartSearch“ recherchieren. Experten können über „DIMDI ClassicSearch“ Recherchen mit komplexen Frage-

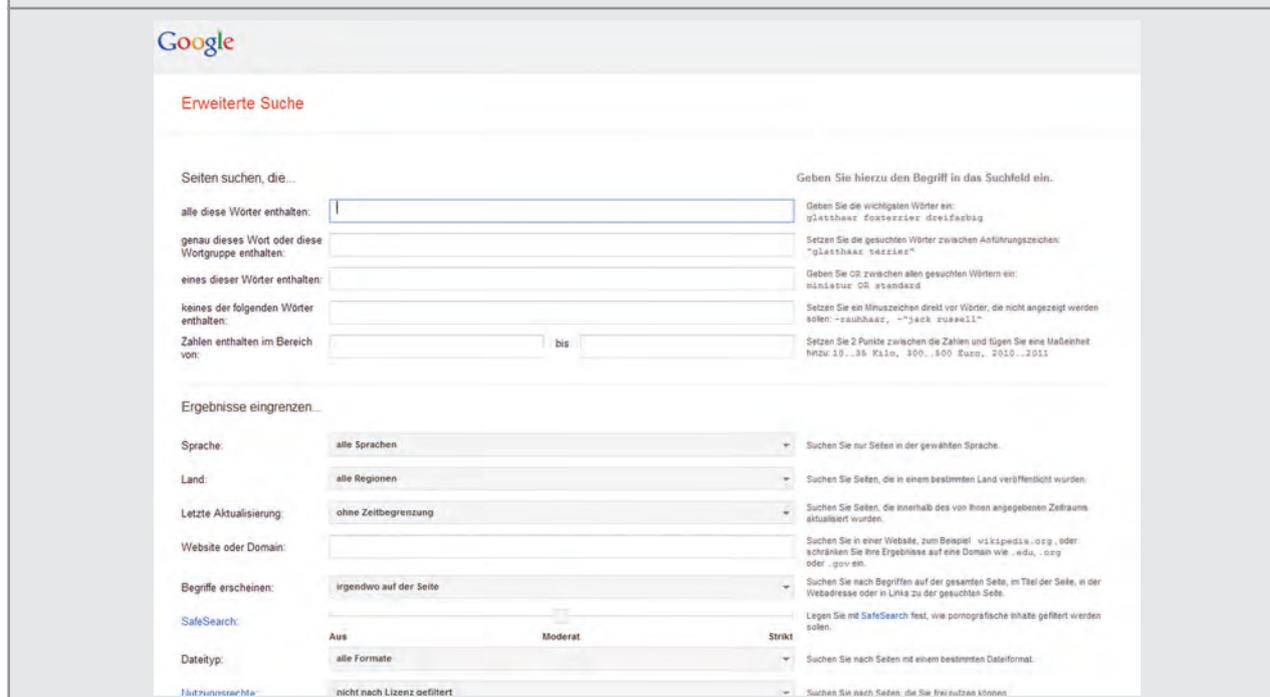
stellungen durchführen. Bei kostenpflichtigen Datenbankprodukten lohnt es sich, die Nutzungs- und Dokumentengebühren verschiedener Datenbankanbieter zu vergleichen.

Die Datenbank EMBASE wird beispielsweise von folgenden Anbietern vertrieben:

- Dialog ([www.dialog.com](http://www.dialog.com))
- ProQuest Dialog ([www.dialog.com/proquestdialog](http://www.dialog.com/proquestdialog))
- Ovid ([www.ovid.com](http://www.ovid.com))
- STN International ([www.stn-international.de](http://www.stn-international.de))
- DIMDI ([www.dimdi.de](http://www.dimdi.de))

Beim DIMDI sind Dokumente aus einigen Datenbanken (MEDLINE, Deutsches Ärzteblatt etc.) zusätzlich zur kostenfreien Recherche auch kostenlos verfügbar (**Abbildung 1**), Dokumente aus EMBASE und BIOSIS sind dagegen kostenpflichtig. Genaue Preisinformationen, Beschreibungen zur Datenbanknutzung oder auch Kurs- und Schulungsangebote bietet das DIMDI auf seiner Website [www.dimdi.de](http://www.dimdi.de) an.

**Abbildung 2: Erweiterte Suche bei Google (Google 2013).**  
 Quelle: [http://www.google.de/advanced\\_search](http://www.google.de/advanced_search) [Abrufdatum: 08.01.2013].



## Suchmaschine Google

Aber nicht nur Datenbanken können als Informationsquelle dienen, sondern auch Suchmaschinen wie Google können, wenn auch eingeschränkt, genutzt werden. Der Unterschied zu Datenbanken besteht bei Suchmaschinen darin, dass sie nicht von eigens dafür zuständigen Experten auf ihren Inhalt und damit ihre Qualität geprüft werden.

Google bietet neben der allseits bekannten einfachen Websuche noch mehr. Wenn auch bei Weitem nicht so differenziert wie Dokumente in gepflegten Datenbanken, können erweiterte Suchen durchgeführt oder Ergebnisse nach gewünschten Formaten gefiltert werden. Anwendung finden zum Beispiel die gängigen logischen Operatoren (UND, ODER, NICHT). Der direkte Einstieg in die erweiterte Suche über die Startseite von Google ist nicht mehr möglich, doch nach dem gewohnten Starten einer einfachen Suchabfrage findet sich am unteren Ende der Seite der Link zur „Erweiterten Suche“ ([www.google.de/advanced\\_search](http://www.google.de/advanced_search); **Abbildung 2**). Mit dieser „Erweiterte Suche“ lässt sich die Suchanfrage erheblich einschränken, was zu einer wesentlich überschaubareren Menge an Ergebnissen führt. Weitere Erläuterungen zu den Suchfunktionen bietet die Google-Hilfe, auf die ebenfalls am Ende jeder Ergebnisseite für eine Suchanfrage hingewiesen wird.

## Fazit

Durch das Internet kann die Masse an vorhandenen Informationen nicht mehr mit einem Blick erfasst und wichtige Dokumente können leicht übersehen werden. Genaue Kenntnisse über das ‚Information Retrieval‘, das Wiederfinden von Informationen, sind für professionelle Recherchen unabdingbar. Die Ware „Information“ wird auf dem Markt als hochsensibles und wertvolles Gut gehandelt, da sie für Forschung und Entwicklung essentiell ist. Deshalb sollten für wichtige Recherchen geschulte Mitarbeiter oder gegebenenfalls extra dafür ausgebildete ‚Information Broker‘ eingesetzt werden, als Basis für gute Recherchen und um Zeit und Kosten zu sparen. Vor allem wenn es um aktuelle, valide Informationen geht, ist die Nutzung von geprüften Datenbanken unbedingt erforderlich. Suchmaschinen, wie beispielsweise Google oder Yahoo, bieten nur wenige Möglichkeiten für gezielte Suchaktionen. Sie sollten daher eher nur zum Erlangen eines ersten Informationsüberblicks eingesetzt werden.

## Weiterführende Informationen

### **DATENBANKANBIETER DIMDI**

**Das DIMDI:** <http://www.dimdi.de/static/de/dimdi/index.htm> (Abrufdatum: 21.01.2012).

**Datenbankrecherche:** <http://www.dimdi.de/static/de/db/index.htm> (Abrufdatum: 21.01.2013)

**Medizinische Literatur und Fakten finden:** <http://www.dimdi.de/static/de/db/basisinfo-db.pdf> (Abrufdatum: 21.01.2013).

### **WEITERE DATENBANKANBIETER**

- Deutsches Ärzteblatt: <http://www.aerzteblatt.de>
- Dialog: <http://www.dialog.com>
- Elsevier B.V.: <http://www.elsevier.com>
- ProQuest Dialog: <http://www.dialog.com/proquestdialog>
- STN International: <http://www.stn-international.de>
- Thomson Reuters: <http://www.ip-science.thomsonreuters.com>
- U.S. National Library of Medicine: <http://www.nlm.nih.gov>

(Abrufdatum aller Datenbankanbieter: 08.01.2013)

### **WEITERE DATENBANKEN**

- German Medical Science (gms): <http://www.egms.de/dynamic/de/index.htm>
- New England Journal of Medicine: <http://www.nejm.org/>
- Registry of Toxic Effects of Chemical Substances (RTECS®), verfügbar über DIMDI: <http://www.dimdi.de>

(Abrufdatum aller Datenbanken: 08.01.2013)

### **INFORMATIONSMITTLUNGSTELLEN**

Öffentlich-rechtliche Informationsvermittlungsstellen (Stand: Mai 2012): <http://www.dimdi.de/static/de/db/service/ivs/ivs1.htm> (Abrufdatum: 08.01.2013).

Privatwirtschaftliche Informationsvermittlungsstellen – Informationsbroker (Stand: August 2012): <http://www.dimdi.de/static/de/db/service/ivs/ivs2.htm> (Abrufdatum: 08.01.2013).

## Literatur

Factsheet Medline (2011): <http://www.nlm.nih.gov/pubs/factsheets/medline.html> (Abrufdatum: 08.01.2013).

Poetzsch E (2005): Naturwissenschaftlich-technische Information: Online, CD-ROM, Internet. 2., völlig neu bearb. Aufl. Berlin: Poetzsch. ISBN: 978-3-938945-00-1.

Poetzsch E (2006): Information Retrieval: Einführung in Grundlagen und Methoden. 5., völlig neu bearb. Aufl. Berlin: Poetzsch. ISBN: 978-3-938945-01-8.

## Kontakt

Carolin Werner, Manuela Franke  
Robert Koch-Institut  
Abteilung für Epidemiologie und  
Gesundheitsmonitoring  
Zentrum für Krebsregisterdaten  
General-Pape-Straße 62-66  
12101 Berlin  
E-Mail: caroline.werner[at]fh-potsdam.de,  
FrankeM[at]rki.de

[RKI]