

# Google Scholar und Pubpharm – was leisten die kostenlosen Suchdienste?

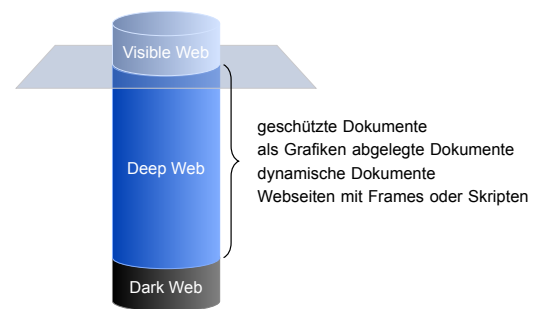
Wolfgang Erdmann

Für Apothekerinnen und Apotheker als Wissensmanager ist es wichtig, gute Suchtools verfügbar zu haben. Google Scholar und PubPharm sind zwei kostenlose Suchdienste. Die folgenden Informationen sollen dabei helfen einzuschätzen, was diese Suchtools leisten und welche möglichen Beschränkungen zu beachten sind.

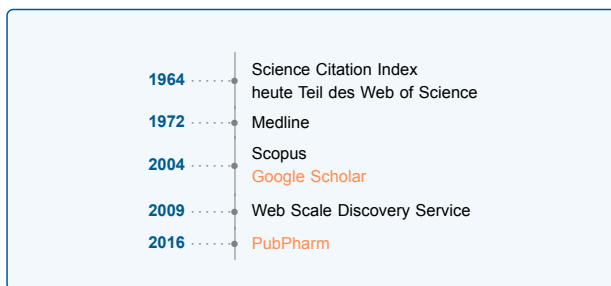
Die wesentlichen Arten von Tools für die Literaturrecherche zeigt die folgende Übersicht:

Literaturdatenbank	Zitations- und Abstractdatenbank	Akademische Suchmaschine	Web Scale Discovery Service
Medline	WoS, Scopus	Google Scholar	PubPharm
Positivliste der Quellen		Internet / Deep Web	versch. Quellen, ein Index
Datenbanktechnologie		Suchmaschinentechnologie	

te werden dem Index zugefügt. Bei der Suche über Google Scholar werden die Trefferlisten durch einen Rankingalgorithmus unter anderem nach der Häufigkeit der Zitationen sortiert ausgegeben. Durch Verträge mit Verlagen enthält der Index auch Inhalte, die dem sogenannten „Deep Web“ zugerechnet werden. Das Deep Web kann man wie folgt darstellen:

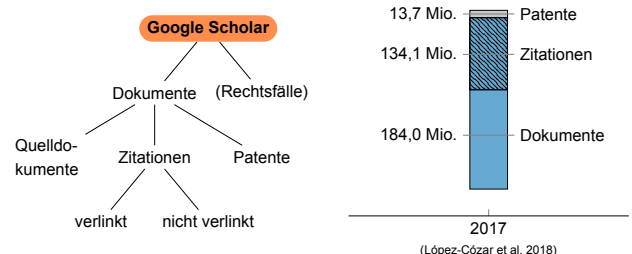


Die zeitliche Entwicklung ist in folgender Zeitleiste dargestellt:



## 1.1 Google Scholar – Fakten und Funktionen

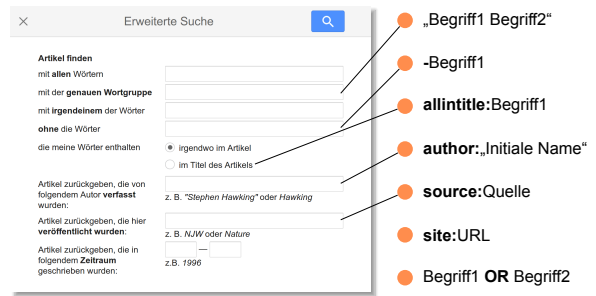
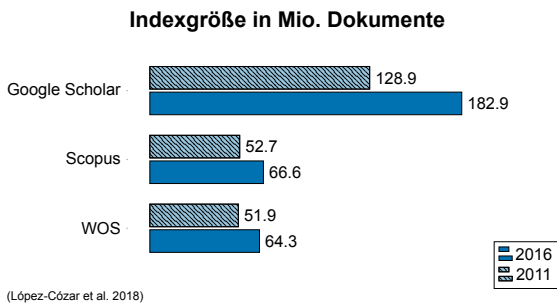
Über den Aufbau und die Größe des Index von Google Scholar gibt die Firma Google nur wenig Informationen. Gemäß einer aktuellen Übersichtsarbeit von López-Cózar et al. setzt sich der Index von Google Scholar wie folgt zusammen:



## 1 Google Scholar

Google Scholar ist eine wissenschaftliche Suchmaschine. Der Index wird durch Algorithmen anhand des Layouts und der Herkunft der Information automatisch erzeugt. Als wissenschaftlich eingestufte Dokumen-

Ein Vergleich mit Scopus und dem Web of Science (WOS) läßt vermuten, das der Index von Google Scholar dreimal größer ist:



Der größte Anteil sind Zeitschriftenartikel, gefolgt von Büchern, Dissertationen und Konferenzbeiträgen. Auf Grund der maschinellen Erstellung des Index gibt es daneben zahlreiche weitere Dokumentenarten. Auf Grund der fehlenden Qualitätskontrolle sind darunter auch Quellen mit minderer wissenschaftlicher Qualität.



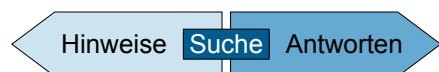
**Ein Tipp:** Da die Anzahl der Zeichen in der einfachen Suchmaske begrenzt ist, kann man bei komplexen Suchanfragen den Operator „OR“ durch den senkrechten Strich (Pipe) auf der Tastatur ersetzen.

Bei einer Recherche sucht man immer in zwei Richtungen. Zum einen möchte man Antworten auf die Recherchefrage finden. Zum anderen schaut man gleichzeitig nach Hinweisen für weitere Quellen. Bei letzterem unterstützt einen Google Scholar durch drei Funktionen in der Trefferanzeige. Zu jedem gefundenen Dokument werden „ähnliche Artikel“ sowie unter „zitiert von“ die Dokumente angeboten, in denen das aktuelle Dokument zitiert wurde. Soweit es verschiedene Versionen eines Dokumentes in Google Scholar gibt, werden diese automatisch zusammengefasst und können unter „alle X Versionen“ angezeigt werden. Hierüber findet man manchmal auch eine kostenlose Version eines sonst kostenpflichtigen Artikels.

López-Cózar et al. geben als wesentliche Quellen für die Dokumente im Index von Google Scholar folgende an:

Quelle
Bibliographische Informationssysteme (PubMed, Europe PubMed Central...)
kommerzielle Verlage (Elsevier, Springer, Wiley, de Gruyter...)
andere akademische Suchmaschinen (Semantic Scholar, Citeseer...)
fachliche Dokumentenserver (arXiv.org)
soziale Plattformen (ResearchGate, Citeseer...)
Googles eigene Plattform (Google Books)
staatliche Einrichtungen (National Institute of Health, NASA...)
Universitäten (Harvard, Universität von Amsterdam...)
internationale Vereinigungen (IEEE - Institute of Electrical and Electronics Engineers)

(López-Cózar et al. 2018)



Beim Einsatz von Google Scholar sollte man folgende Einschränkungen kennen und berücksichtigen:

In Google Scholar kann man neben der einfachen Suche auch eine erweiterte Suchmaske nutzen. Der folgenden Übersicht kann man entnehmen mit welchen Operatoren man im einfachen Suchschlitz die gleichen Ergebnisse wie mit den entsprechenden Feldern in der erweiterten Suche erzielt:

1. Der Suchschlitz fasst nur eine begrenzte Anzahl an Zeichen (255). Dies kann bei sehr komplexen Suchanfragen limitierend sein.
2. In der Trefferliste werden nur die ersten 1000 Treffer angezeigt. Dies kann erheblichen Einfluss auf die Recall-Rate haben. Die Suchergebnisse

- können nicht in Gänze heruntergeladen werden.
3. Es gibt keinen Operator für die Wortstammsuche, dafür aber eine automatische Wortstammsuche.
  4. Die Änderung der Reihenfolge von mit OR verknüpften Suchbegriffen ergibt unterschiedliche Treffermengen. Bei der Google-Suche werden weiter vorne stehende Begriffe als wichtiger gewertet.
  5. Die Verschachtelung logischer Operatoren ist nur auf einer Ebene möglich; eine UND-Verknüpfung der Begriffe geschieht automatisch durch Leerzeichen dazwischen; ODER-Verknüpfungen mit „OR“ sind nur auf der selben Ebene möglich und müssen in Klammern gesetzt werden.
  6. Es gibt kein kontrolliertes Vokabular (Thesaurus).
  7. In der erweiterten Suche gibt es je Suchfunktion nur ein Eingabefenster; komplexe Suchanfragen müssen daher in einem externen Texteditor formuliert und in das einfache Suchfenster kopiert werden.
  8. Bei PDF-Dateien über 5 MB Größe wird der Volltext nicht indiziert.
  9. Es gelten die gleichen Datenschutzbestimmungen wie auch bei der Suche mit Google.

## 1.2 Studien über Google Scholar

Zur Charakterisierung von Literaturdatenbanken werden oft folgende Messgrößen herangezogen:

Fragestellung: 250 relevante Texte bekannt  
 Datenbank: enthält 100 relevante Texte  
 Suchergebnis: 75 Treffer – 50 relevant, 25 irrelevant

$$\text{Recall} = \frac{\text{Anzahl der gefundenen relevanten Dokumente}}{\text{Anzahl aller relevanten Dokumente in der DB}} = \frac{50}{100} = 0,5$$

$$\text{Precision} = \frac{\text{Anzahl der gefundenen relevanten Dokumente}}{\text{Anzahl aller gefundenen Dokumente}} = \frac{50}{75} = 0,66$$

$$\text{Coverage} = \frac{\text{Anzahl der enthaltenen relevanten Dokumente}}{\text{Anzahl aller bekannten relevanten Dokumente}} = \frac{100}{250} = 0,4$$

**Bramer et al.** haben 2016 in einer Studie die Suchergebnisse von Google Scholar, Medline und Embase verglichen. Gesucht wurde ein Goldstandard von 4795 Referenzen aus 120 Systematic Reviews. Das Vorgehen war wie folgt: Ein Suchspezialist hat mit angepassten Suchanfragen in den verschiedensten Datenbanken Arbeiten für Systematic Reviews gesucht. Die Ergebnisse von Embase, Google Scholar (die ersten 1000 Treffer) und Medline wurden in einer Endnote-

Datei gesondert erfasst. Die Anfragenden haben dann aus dem Gesamtergebnis die Studien für die Reviews nach den Ein- und Ausschlusskriterien ausgesucht. Anschließend wurde durch gezielte Suche in den drei Datenbanken festgestellt, welche Studien aus dem Goldstandard zwar im jeweiligen Index enthalten, mit der Suchstrategie aber nicht gefunden wurden. Für die drei Suchdienste wurden Coverage, Recall und Precision ausgerechnet.

**Ergebnis:** Die Coverage von Google Scholar lag bei 97,2%, von Embase bei 97,5% und von Medline bei 92,2%. Der Recall von Medline plus Embase war mit 81,6% am größten. Google Scholar hatte einen Recall von 72,8%, aber innerhalb der ersten 1000 Treffer waren nur 46,4%; die Precision von GS bei den ersten 1000 Treffern war 1,9%. Die Precision für Embase lag bei 2% und für Medline bei 2,8%.

**Fazit:** keine der drei Datenbanken kann als alleinige Quelle für die Suche nach Arbeiten für Systematic Reviews genutzt werden. Auch bei der Kombination von Medline und Embase gab es für die Arbeiten aus einzelnen Systematic Reviews nur einen Recall von weniger als 50%. Wenn man nur die ersten 1000 Treffer zu Grunde legt ist die Precision von Google Scholar vergleichbar mit der der anderen Datenbanken.

**Cicccone und Vickery** verglichen 2015 die Ergebnisse typischer Suchanfragen von Bibliotheksnutzern bei Nutzung von zwei Discovery Services (Summon und EBSCO Discovery Service) und Google Scholar anhand einer Stichprobe von 183 Suchanfragen aus einem Log-File. Bei thematischen Suchanfragen wurde die Zahl der relevanten Ergebnisse innerhalb der ersten zehn Treffer ausgewertet. Der Grund hierfür ist, dass sich laut Studien Nutzer von Bibliotheksdatenbanken selten mehr als die erste Trefferseite ansehen. ¼ der Nutzer suchen bekannte Artikel, ¾ machen thematische Suchen mit häufig nur zwei bis 3 Schlagwörtern. Bei der Suche nach bekannten Artikeln wurde untersucht, ob der Artikel gefunden wurde.

**Ergebnis:** Es gab bei den Discovery Services keinen signifikanten Unterschied der Ergebnisse sowohl bei der Suche nach bekannten Artikeln als auch bei der thematischen Suche. Zwischen beiden Discovery Services und Google Scholar gab es bei der Suche nach bekannten Artikeln ebenfalls keinen Unterschied. Google Scholar übertraf jedoch mit durchschnittlich 5,68 relevanten Treffern innerhalb der ersten zehn Treffer sowohl Summon (4,76 Treffer) als auch den EBSCO

Discovery Service (4,83 Treffer). Eine Schwäche der Studie war die geringe Anzahl der Suchen nach bekannten Artikeln (44 von 183 = 24%).

### 1.3 Google Scholar Zusatzangebote

**Library Links (Bibliothekslinks)** ist seit 2005 ein Service bei dem man unter „Einstellungen – Bibliothekslinks“ aus einer Liste mit Google Scholar kooperierender Bibliotheken maximal fünf auswählen kann. Für die Dokumente aus der Trefferliste wird dann geprüft, ob diese im Bestand der ausgewählten Bibliothek sind. Ist dies der Fall, kann man per Link aus der Trefferliste direkt den Eintrag im Katalog der Bibliothek aufrufen.

**Google Scholar Alerts (Benachrichtigungen)** ermöglicht seit 2010 eine Suchanfrage abzuspeichern. Per E-Mail werden dann zu dieser Suche passende neu indexierte Dokumente gemeldet.

**Google Scholar Citations (GSC)** existiert seit 2011 als Plattform für akademische Profile von Autoren und enthält deren Veröffentlichungen und Zitationen. Der Nutzen für den Autor ist, dass er immer eine aktuelle Liste seiner Publikationen hat. Ein Nutzen für andere ist, dass man über GSC nach Experten für ein Fachgebiet suchen kann.

**Google Scholar Metrics (GSM)** gibt es seit 2012 und ist ein Ranking von wissenschaftlichen Publikationen nach bibliometrischen Indikatoren (auf Basis von Zitationsdaten der in den Zeitschriften veröffentlichten Artikel).

## 2 PubPharm

PubPharm ist eine vom Fachinformationsdienst Pharmazie (FID) der Universität Braunschweig entwickelte Suchmaschine für pharmazeutische Fachliteratur.

### 2.1 PubPharm – Fakten und Funktionen

PubPharm basiert auf K10Plus-Zentral, einer Suchmaschine für bibliographische Daten, welche die Verbundzentrale Göttingen (VZG) in Kooperation mit Bibliotheksservice-Zentrum Baden-Württemberg (BSZ)

betreibt. Die Suchmaschine hat keine eigene Oberfläche. Durch Filter wird der Zugriff bei PubPharm auf die pharmaziespezifischen Einträge beschränkt. Der Index von K10Plus-Zentral sowie die Selektion für PubPharm setzen sich wie folgt zusammen:

Medien	Anteil (ca.)
Medline	29.092.000
Artikel	26.344.000
gedruckte Bücher	661.000
eBooks	242.000
Sonstige (Dissertationen)	246.000
<b>Gesamt</b>	<b>56.585.000</b>

K10plus-Zentral des GBV	
Daten von ca. 530 Bibliotheken	
+ Medline	
+ Daten von Verlagen	
+ elektr. Zeitschriftenbibliothek	
über 207 Mio. Datensätze	
davon 49,9 Mio. Artikel	
aus > 24.000 Zeitschriften	

Stand Januar 2019

Bei der Suche über den einfachen Suchschlitz werden automatisch passende MeSH-Begriffe aus Medline bzw. Stoffnamen aus der Substanzdatenbank DrugBank vorgeschlagen. Die Sortierung der Trefferliste ist nach Jahr oder Relevanz möglich. Durch eine automatische Verfügbarkeitsprüfung wird angezeigt, ob ein Dokument frei verfügbar oder kostenpflichtig abrufbar ist bzw. über Lizenzen des FID berechtigten Nutzern zur Verfügung steht. Die wichtigsten Operatoren für die Suche in PubPharm sind:

Suchfunktionen	Operator	Funktion
Suchwortvorschläge	AND	Begriffe einschließen
Sortierung nach Jahr oder Relevanz	OR	Suche erweitern
Verfügbarkeitsprüfung	NOT oder -	Begriffe ausschließen
	* Stern	Wortstammsuche

Bei der Verwendung von PubPharm bestehen derzeit noch folgende Einschränkungen, welche im Rahmen der ständigen Weiterentwicklung behoben werden sollen:

1. Die Trefferlisten enthalten z. T. Dubletten.
2. Die Verarbeitung komplexer Suchanfragen mit verschachtelten Operatoren über den Suchschlitz ist noch nicht intuitiv.
3. Bei den Suchfiltern werden aus Performanzgründen lediglich die ersten 20 Einträge angezeigt.
4. Durch die unterschiedliche Erfassungsqualität

der Datenquellen fehlen teilweise für die Filter benötigte Angaben im Datensatz. Das bewirkt, dass man z. B. bei Filterung nach Dokumenten in deutscher Sprache nur einen Teil der in der Datenbank vorhandenen Datensätze angezeigt bekommt.

- Der Index des GBV weist in den 60er bis 80er Jahren Lücken auf.





## 2.2 PubPharm – innovative Suchtools

**Ähnlichkeitssuche:** Bei der Suche nach einem Arzneistoff werden Kontext-verwandte Substanzen vorgeschlagen. Diese Vorschlagsfunktion basiert auf Deep-Learning-Techniken.





Die **Struktursuche** bietet insbesondere für die Pharmazeutische Chemie einen Mehrwert gegenüber der textbasierten Suche. Mit einem Editor können chemische Strukturen eingegeben werden oder aber es wird nach der (englischen) Stoffbezeichnung gesucht. Bei Eingabe der Struktur versucht PubPharm diese einer Stoffbezeichnung zuzuordnen. Für erkannte bzw. vorgegebene Stoffe kann nach Substrukturen oder strukturell ähnlichen Substanzen gesucht werden.

## 3 Vorteile von Google Scholar und PubPharm





Vorteile für **Gelegenheitssucher**

-   einfache, intuitive Suchoberfläche
-  hohe Quote relevanter Treffer – Qualitätskontrolle!
-  Schwerpunkt Nachweis pharmazeutischer Literatur




Vorteile für **regelmäßige Sucher**

-  Suche nach ähnlichen Wirkstoffen
-  Suchvorschläge aus MeSH und DrugBank
-  Suchhinweise durch „ähnliche Artikel“ und „zitiert durch“
-  Alerting-Dienst

Vorteile für **wissenschaftliche Sucher**

-  Struktursuche
-  Nachweis von Dissertationen und Büchern
-  als Ergänzung zum Nachweis Grauer Literatur
-  als Alternative für die Zitationssuche

## 4 Quellenangaben

-  Bramer, Wichor M., Dean Giustini und Bianca M. R. Kramer (März 2016). „Comparing the Coverage, Recall, and Precision of Searches for 120 Systematic Reviews in Embase, MEDLINE, and Google Scholar: A Prospective Study“. In: *Systematic Reviews* 5.1, S. 39. ISSN: 2046-4053. DOI: 10.1186/s13643-016-0215-7.
-  Ciccone, Karen und John Vickery (März 2015). „Summon, EBSCO Discovery Service, and Google Scholar: A Comparison of Search Performance Using User Queries“. en. In: *Evidence Based Library and Information Practice* 10.1, S. 34–49. ISSN: 1715-720X. DOI: 10.18438/B8666Q.
-  Delgado López-Cózar, Emilio Delgado, Enrique Orduna-Malea und Alberto Martín-Martín (Juni 2018). „Google Scholar as a Data Source for Research Assessment“. In: *arXiv:1806.04435 [cs]*. arXiv: 1806.04435 [cs].

## 5 Extratipp

So kann PubPharm als Auswahlmöglichkeit zum Suchfenster des Browsers Firefox zugefügt werden:

- PubPharm-Startseite ([www.pubpharm.de](http://www.pubpharm.de)) mit Firefox aufrufen
- Lupensymbol mit grünem Pluszeichen links neben dem Suchschlitz anklicken
- „PubPharm Search hinzufügen“ anklicken