

Warum stritten wir je um Discovery und OPAC?

# „Boutique“-Recommender-Systeme als aktuelles Aufgabengebiet für Digitale Bibliotheken

David Zellhöfer

» Nach einigen spannenden Jahren in der Wissenschaft wunderte ich mich nach meinem Wechsel in die Welt der wissenschaftlichen Bibliotheken, dass ich im Jahr 2014 noch die letzten Ausläufer der Diskussion um die Daseinsberechtigung von OPAC- und Discovery-Systemen erleben durfte, während man in der Wissenschaft, die Entwicklung vom Boole'schen Retrievalmodell<sup>1</sup> hin zu besser im Umgang mit natürlicher Sprache geeigneten statistisch-algebraisch motivierten Modellen wie  $tf*idf$ <sup>2</sup>, die in den 1970/80er Jahren stattfand, als unausweichlichen evolutionären Prozess betrachtete. Die damals im Bereich der wissenschaftlichen Bibliotheken geführte Diskussion stellte OPAC und Discovery als Antagonisten dar<sup>3</sup>, wobei sich die letztgenannten wesentlich näher an den Suchgewohnheiten der üblicherweise im WWW recherchierenden Nutzerinnen und Nutzer orientierten und sich deshalb berechtigterweise als nutzerorientierter Ansatz etablierten, während sich parallel – und durch die Bibliotheken scheinbar unbemerkt – Recommender-Systeme und maschinelles Lernen als neue Forschungsgebiete spätestens seit den 1990/2000er Jahren herausbildeten<sup>4</sup>.

## Was bedeutet Nutzerorientierung mit Hinblick auf Discovery?

In einem mittlerweile gereiften Beitrag stellt Block<sup>5</sup> zwei Hauptaspekte moderner Discovery-Systeme heraus: die große Abdeckung an Bibliotheksbeständen während des Suchprozesses und die Bereitstellung eines modernen User-Interfaces (UI).

Block weist darauf hin, dass ein gutes Discovery-System Volltextinhalte, den Bibliothekskatalog, Fachdatenbanken, E-Books und institutionelle Repositorien einbinden soll. Die Modernität sieht er primär durch

### Abstract

*Der Artikel stellt ein aktuelles Tätigkeitsfeld für Bibliotheken vor: die Gestaltung von Recommender-Systemen als nutzerorientierte Überwindung des Query-Response-Paradigmas von Discovery-Systemen. Basierend auf der These, dass Nutzende bei der Interaktion mit Beständen Unterstützung benötigen, werden Defizite gängiger Discovery-Systeme skizziert. Es wird herausgearbeitet, dass die Einflussmöglichkeiten auf die Gestaltung von Discovery-Systemen mit Hinblick auf Usability und die Unterstützung moderner Suchstrategien aktuell begrenzt ist. Für innovative Bibliotheken sollte es deshalb künftig eine Priorität sein, sich auf Methoden aus dem Bereich der Recommender-Systeme zu stützen, um nicht den Anschluss an die klassische WWW-Suche und die damit verbundenen Nutzungserwartungen zu verlieren.*

*The article presents an upcoming field of work for libraries: the design of recommender systems in order to overcome the classic query response paradigm. Based on the observation that users need assistance during the interaction with digital libraries, shortcomings of current discovery systems are outlined. Furthermore, the article emphasizes the fact that adaption opportunities of discovery systems regarding usability needs or the support of modern information seeking strategies is currently limited. Hence, the article argues that innovative libraries have to make an effort to exploit methods from the field of recommender systems in order to stay on par with modern WWW search engines and current user expectations.*

eine zeitgemäße UI begründet. Block erkennt dabei richtig, dass erst die Kombination aus Auffindbarkeit von Materialien in Kombination mit einem bedienbaren Werkzeug, der UI, ein effizient und effektiv nutzbares Werkzeug im Sinne guter Usability ergibt.

Die Zuschreibung des Modernitätsaspekts zu Discovery-Systemen hat auch 2022 laut Holländer<sup>6</sup> seine Berechtigung. Holländer führt weiter aus, dass erfolgreiches Discovery (unter Rückgriff auf Block<sup>7</sup>) auf vier

1 Codd, E. F. (1970). A relational model of data for large shared data banks. Commun. ACM, 13(6), 377-387. bzw. Waller, W. G., & Kraft, D. H. (1979). A Mathematical Model of a Weighted Boolean Retrieval System. Information Processing and Management, 15(5), 235-245.

2 Spärck Jones, K. (1972). A Statistical Interpretation of Term Specificity and Its Application in Retrieval. Journal of Documentation, 28(1), 11-21.

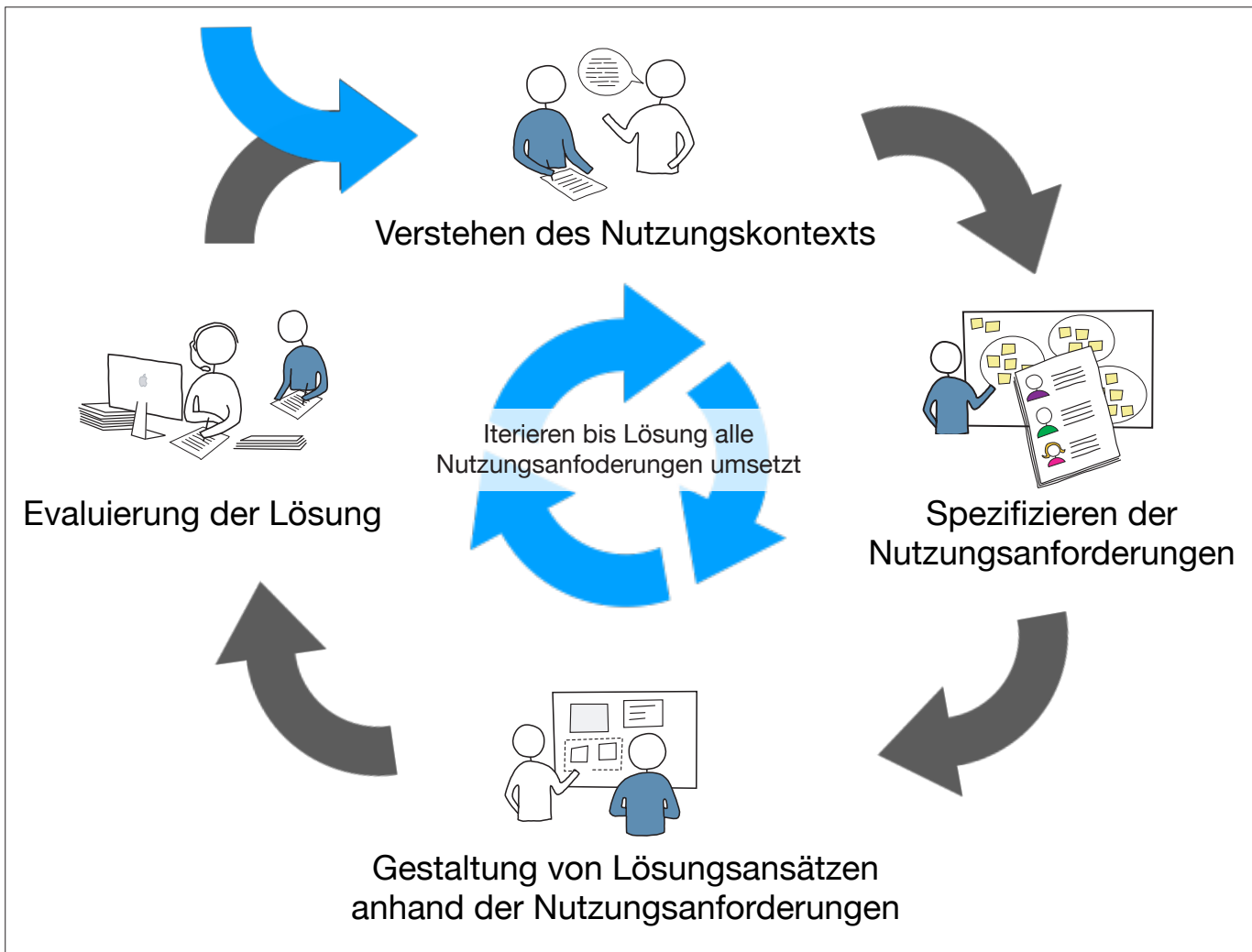
3 Eine Einschätzung, die sich theoretisch nicht aus den unterschiedlichen Retrieval-Modellen herleiten lässt, wenn man die Relevanzbewertung der  $tf*idf$ -basierten Systeme als Wahrscheinlichkeit interpretiert. Aus Platzgründen kann dieser Gedanke hier jedoch nicht ausgeführt werden.

4 Aggarwal, C. C. (2016). Recommender Systems: The Textbook. Springer.

5 Block, N. (2015). Das Bibliothekssystem neu überdenken. b.i.t. online, 18(2), 164-167.

6 Holländer, S. (2022). Mehr Kundennähe wagen! b.i.t. online, 25(1), 28-33.

7 Block, N. (2015).



**Abbildung 1:**  
Nutzerorientierter, iterativer UCD-Prozess

Pfeilern ruht: der modernen UI, einer umfassenden Inhaltsintegration, einem Relevanzranking und Optionen zur Anpassung der Software. Neben dem starken Fokus auf die Inhaltsabdeckung und seiner Forderung nach der aktiven Einbeziehung von Nutzendenfeedback in die Fortentwicklung von Discovery-Systemen, benennt Holländer den Umstand, dass Bibliotheken nicht mehr das Informationsmonopol innehaben, sondern mit modernen Suchmaschinen des WWW konkurrieren<sup>8</sup> und damit riskieren, Nutzende zu verlieren, wenn sie sich nicht an deren Bedürfnissen orientieren.

Aus Sicht der Bibliotheken ist es gut nachvollziehbar, dass die Abdeckung einer Vielzahl an Beständen eine wesentliche Rolle bei der Bewertung von Discovery-Systemen spielt, da sie theoretisch die potentielle Retrievability<sup>9</sup> von Beständen erhöht. Problematisch bei dieser Annahme ist jedoch, dass Nutzende nicht im gleichen Maße wie Bibliothekarinnen und Bibliothekare um die umfassenden Bestände einer Bibliothek

wissen, sondern diese zumeist erst kennenlernen müssen und dabei bestmöglich unterstützt werden sollten.

Unter diesem Aspekt lohnt es sich, einen Blick auf die Fortentwicklungsmöglichkeiten für Bibliotheken im Bereich der unterstützenden Informationssuche zu werfen.

### Zur Nutzerorientiertheit von Discovery-Systemen

In seinem Artikel kritisiert Holländer<sup>10</sup>, dass Bibliotheksnutzerinnen und -nutzer keine Lobby haben und bei schlechter Usability abwandern. Während der letztgenannte Punkt unstrittig ist, überrascht der erstgenannte Punkt, da er suggeriert, dass ein Defizit an Wissen über den Umgang von Nutzenden mit den Herausforderungen der Informationssuche existiere. Dem können diverse Studien entgegengehalten werden, welche teils schon weit zurückliegen und einen expliziten Nutzerstandpunkt einnehmen. Die Verfüg-

<sup>8</sup> Vgl. Holländer, S. (2022); S. 30

<sup>9</sup> Azzopardi, L., & Vinay, V. (2008). Retrievability: An Evaluation Measure for Higher Order Information Access Tasks. Proceedings of the 17th ACM Conference on Information and Knowledge Management, 561-570.

<sup>10</sup> Holländer, S. (2022). Mehr Kundennähe wagen! b.i.t. online, 25(1), 28-33.

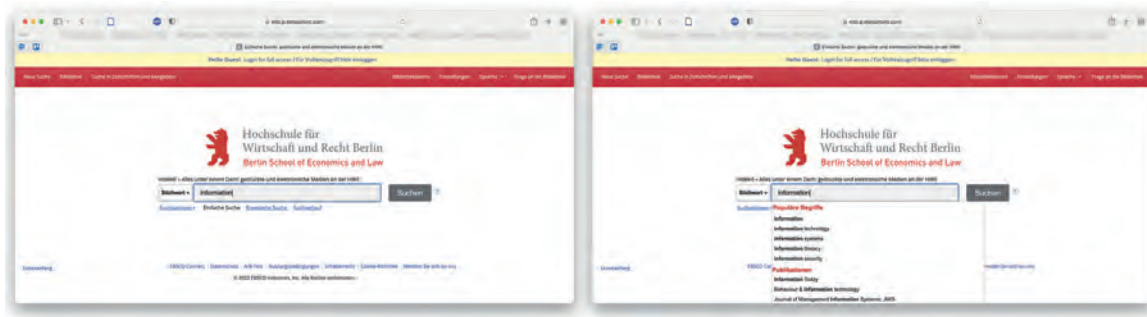


Abbildung 2: Typische Discovery-Einstiegsseite am Beispiel der HWR Berlin (EBSCO EDS) inkl. Schlagwort-basierten Vorschlägen (rechts)

barkeit an Studien beginnt nicht erst bei Kuhlthau<sup>11</sup>, sondern erstreckt sich über immer noch aktuelle und frei verfügbare Kompendien wie das Werk von Hearst<sup>12</sup> bis hin zu konkreten Usability-Studien von Discovery-Systemen<sup>13,14</sup>. Studien mit Nutzenden, teils sogar mit explizit benannten Nutzendengruppen, sind noch immer ein Thema im Bereich der Bibliotheks- und Informationswissenschaften, wie aktuelle Abschlussarbeiten<sup>15</sup> zeigen. Offenbar besteht hier eine Hürde, dass diese Erkenntnisse auch in der Praxis umgesetzt werden. Welche Hürde dafür ursächlich sein könnte, soll im Folgenden ergründet werden.

Ein weiteres Desiderat, für welches der Artikel von Holländer<sup>16</sup> stellvertretend steht, ist die Forderung, dass Bibliotheken Usability-Studien ihrer angebotenen Systeme kontinuierlich durchführen müssen, um nutzerorientierte Dienste anbieten zu können. Abbildung 1 illustriert diesen nutzerorientierten, iterativen Entwicklungsprozess: den User-Centered-Design-Prozess (UCD).

Charakteristisch für den UCD-Prozess ist die stetig wiederkehrende Erhebung von Nutzendenfeedback, welches in der Anpassung des betrachteten Systems (siehe „Gestaltung von Lösungsansätzen anhand der Nutzungsanforderungen“) resultiert. Es ist erfreulich, dass diese Ideen mittlerweile in den Entwurf neuer Systeme, wie z.B. dem Handschriftenportal<sup>17</sup>, integriert sind bzw. in der Weiterentwicklung umgesetzt werden<sup>18,19</sup>.

Die drei zuvor genannten Beispiele eint, dass die beteiligten Bibliotheken die Entwicklung der Dienste aktiv gestalten. Im Falle des Handschriftenportals

heißt dies, dass zwei der vier Partnereinrichtungen, die Staatsbibliothek zu Berlin und die Universitätsbibliothek Leipzig, gemeinsam und agil ein neues System entwickeln und direkt Anforderungen der Nutzenden umsetzen können.

Überträgt man dies auf den aktuellen Einsatz von Discovery-Systemen an Bibliotheken wird schnell deutlich, dass die nötige Gestaltungsmacht, Änderungen – zumindest beim Einsatz von proprietären Systemlösungen – umzusetzen, nicht bei den Bibliotheken liegt. Bibliotheken können in diesem Fall zwar Defizite in der Bedienbarkeit ihrer Discovery-Systeme erkennen, wenn Sie Usability-Studien durchführen, diese aber kaum bis gar nicht abstellen, da sie auf den Kooperationswillen des proprietären Softwareanbieters angewiesen sind. Letztendlich entscheidet dieser folglich, wie er Modernität und Nutzerorientierung definiert – so wie es auch bei der Abdeckung der Bestände durch die Zusammenstellung der bereitgehaltenen Indizes oder das unterstützte Ingest/Update-Intervall der Bibliothek der Fall ist. Im Extremfall kann eine negative Bewertung seitens der Nutzenden dann nur noch zur Abkehr vom bisherigen Discovery-Dienstleister führen, um Abwanderungen zu vermeiden, wobei unklar ist, ob gewünschte Features seitens anderer Bieter angeboten werden.

### Zur Modernität von Discovery-Systemen

Grundlegende, explorative Suchstrategien werden mittlerweile in allen Discovery-Systemen, egal ob proprietär oder Open Source, in der Regel durch eine facettrierte Suche anhand verschiedener Krite-

11 Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5), 361-371.

12 Hearst, M. A. (2009). *Search user interfaces*. Cambridge Univ. Press. <http://searchuserinterfaces.com>

13 Fagan, J. C., Mandernach, M. A., Nelson, C. S., Paulo, J. R., & Saunders, G. (2008). Usability Test Results for a Discovery Tool in an Academic Library. *Information Technology and Libraries*, 31(1), 83-112.

14 Niu, X., Zhang, T., & Chen, H. (2014). Study of User Search Activities with Two Discovery Tools at an Academic Library. *International Journal of Human-Computer Interaction*, 30(5), 422-433.

15 Bulmahn, M. (2022). *Das Informationsverhalten von Studierenden im digitalen Zeitalter – Eine Handlungsempfehlung für Bibliotheken* [Bachelor Thesis].

16 Holländer, S. (2022).

17 <https://handschriftenportal.de/kommunikationsstrategie/>

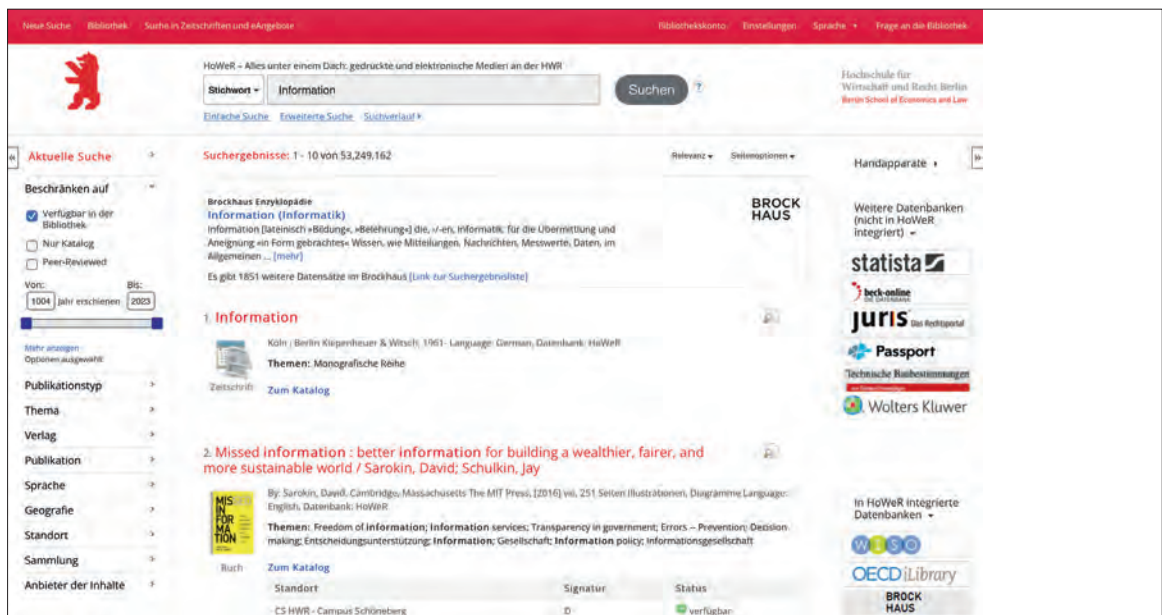
18 <https://etahoffmann.staatsbibliothek-berlin.de/nutzungsanalyse/>

19 [https://blog.sbb.berlin/usability\\_test/](https://blog.sbb.berlin/usability_test/)

Abbildung 3:  
Typische Ein-  
stiegsseite bei  
Google bzw. der  
TU Berlin  
(Ex Libris Primo)



Abbildung 4:  
Ergebnisliste  
der Anfrage  
„Information“  
im Discovery-  
System der  
HWR Berlin  
(EBSCO EDS,  
iPadOS)



rien wie Medientyp, Erscheinungsdatum o.ä. implementiert.

Typischerweise bieten aktuelle Discovery-System-Lösungen Autovervollständigungs- bzw. Schlagwortbasierte Vorschlagsdienste für Nutzende an, welche sich an den Angeboten großer WWW-Suchmaschinen (siehe Abbildung 3, links) orientieren, aber in ihrer Leistungsfähigkeit variieren (siehe Abbildung 2, rechts bzw. Abbildung 3, rechts).

Die Discovery-Systeme folgen dem Query-Response-Paradigma und präsentieren ihre Ergebnisse in einer nach Relevanz geordneten Liste, wobei die explorativen Suchwerkzeuge zumeist entsprechend der westlichen Leserichtung am linken Bildschirmrand platziert werden. Verweise auf weitere Datenbanken werden am rechten Rand positioniert, während Treffer in Nachschlagewerken häufig als Teil der Suchergebnisse mittig eingeblendet werden (siehe Abbildung 4). Der Bildschirmaufbau folgt dabei typischen Best-Practices aus dem Bereich von Web-Suchoberflächen, wie sie sich auch in Web-Shops oder vergleichbaren Websites finden.

Die Bestimmung der Relevanz einzelner Dokumente basiert auf Inhaltsmerkmalen des nachgewiesenen Bestands (z.B. dem Volltext) sowie Feld-basierten Metadaten (z.B. Titel, Autorenschaft, Ressourcentyp o.ä.). Bibliotheken haben in der Regel (auch bei proprietären Anbietern) die Möglichkeit, auf die Relevanzbewertung Einfluss zu nehmen, indem sie die Gewichtung verschiedener Felder bei der Relevanzbeurteilung unterschiedlich gewichten<sup>20,21</sup>.

Die sich daraus ergebenden Steuerungsmöglichkeiten, um die Ergebnispräsentation mit Hinblick auf Anforderungen verschiedener Nutzungsgruppen anzupassen, sind folglich sehr eingeschränkt und zielen primär auf die Unterstützung gerichteter Suchstrategien ab; optimalerweise in einer Art und Weise, dass die wahrgenommene Wichtigkeit der verschiedenen Felder durch die Systembibliothekarinnen und -bibliothekare denen der Nutzenden entspricht. Der starke Fokus auf diese verschiedenen Felder im Rahmen der Relevanzbewertung überrascht nicht, wenn man die zugrundeliegenden Algorithmen bzw. deren Implementierung betrachtet.

20 [https://knowledge.exlibrisgroup.com/Primo/Product\\_Documentation/020Primo\\_VE/Primo\\_VE\\_\(English\)/040Search\\_Configurations/Configuring\\_the\\_Ranking\\_of\\_Search\\_Results\\_in\\_Prime\\_VE](https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_(English)/040Search_Configurations/Configuring_the_Ranking_of_Search_Results_in_Prime_VE)

21 [https://connect.ebsco.com/s/article/How-is-relevance-ranking-determined-in-EBSCO-Discovery-Service-EDS?language=en\\_US](https://connect.ebsco.com/s/article/How-is-relevance-ranking-determined-in-EBSCO-Discovery-Service-EDS?language=en_US)

Moderne Volltextsuchmaschinen basieren algorithmisch auf dem bekannten tf\*idf-Modell bzw. technisch direkt oder indirekt zumeist auf der erstmalig in 1997 erschienenen Lucene<sup>22</sup>-Softwarebibliothek, welche grundlegende Funktionen eines Retrievalsystems bietet. So basieren beispielsweise die beiden Markt-dominierenden Suchmaschinen Solr und Elasticsearch auf Lucene.

ExLibris Primo VE nutzt wiederum, ebenso wie VuFind und finc, Solr als internes Framework zur Volltextsuche und -indizierung<sup>23,24,25</sup>. Für die Repository-Suche in Alma ist der Einsatz von Lucene gesichert<sup>26</sup>. Eine abschließende Beurteilung für EBSCO EDS konnte nicht erfolgen. Lucene indiziert Dokumente intern konzeptionell als sogenannte Documents, welche wiederum für eine Sammlung von Fields dienen.

Diese Fields können dann während der Anfrageverarbeitung unterschiedlich gewichtet werden<sup>27</sup>.

Die konkreten Steuerungsmöglichkeiten aktuell eingesetzter Discovery-Systeme richten sich demnach nicht nach den eventuellen Bedarfen der Nutzenden oder der Bibliotheken, sondern allein nach den technischen Möglichkeiten des zumeist zugrundeliegenden Retrievalsystems Lucene.

### Von Query-Response hin zum „Boutique“-Recommender-System

Sowohl Discovery-Systeme als auch WWW-Suchmaschinen müssen damit umgehen, dass Nutzerinnen und Nutzer nur wenig Schlagworte nutzen, um ihren Informationsbedarf zu spezifizieren. In der Regel greifen sie dabei auf zwei bis maximal fünf Begriffe

22 <http://www.h-online.com/open/news/item/Ten-years-of-the-Lucene-search-engine-at-Apache-1350761.html>

23 [https://dachela.org/pub/DACHELA/Tagung2019/CDI\\_Provider\\_zone\\_de.pdf](https://dachela.org/pub/DACHELA/Tagung2019/CDI_Provider_zone_de.pdf)

24 <https://vufind.org/vufind/features.html>

25 <https://finc.info/de/datenquellen>

26 [https://knowledge.exlibrisgroup.com/Alma/Knowledge\\_Articles/Repository\\_Search\\_-\\_How\\_are\\_Results\\_Sorted\\_\(Ranked\)%3F](https://knowledge.exlibrisgroup.com/Alma/Knowledge_Articles/Repository_Search_-_How_are_Results_Sorted_(Ranked)%3F)

27 [https://lucene.apache.org/core/9\\_2\\_0/core/index.html](https://lucene.apache.org/core/9_2_0/core/index.html)





## Bibliotheksservice der besonderen Art!



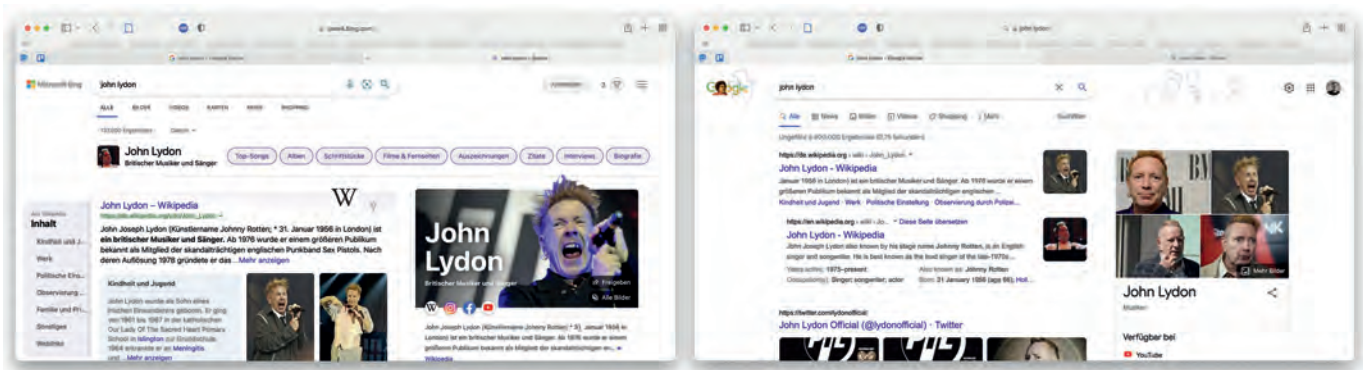



Erleichtern Sie Ihren Erwerbungsalltag und nutzen Sie die breite Vielfalt des DreierService!

[www.dietmardreier.de](http://www.dietmardreier.de)

**Dietmar Dreier** Wissenschaftliche Versandbuchhandlung GmbH  
International Library Suppliers

**Wir freuen uns auf Ihre Kontaktaufnahme:**  
[www.dietmardreier.de](http://www.dietmardreier.de) | Fon: 0049 (0)2065 - 77 55 10 | E-mail: [info@dietmardreier.de](mailto:info@dietmardreier.de)



**Abbildung 5:** *Ergebnislisten bei Bing und Google für die Anfrage „john lydon“ (v.l.n.r.)*

zurück<sup>28,29</sup>. Um diesem Anspruch gerecht zu werden, begannen WWW-Suchmaschinen bereits frühzeitig damit, auf Logdateien zurückzugreifen, um durch verwandte Anfragen das Informationsbedürfnis besser einschätzen zu können und die Anfrage entsprechend zu expandieren (Query Expansion). Typischerweise werden außerdem weitere Quellen wie Standortinformationen herangezogen, um die Anfrage weiter zu präzisieren (z.B. wird die Anfrage „italienisches Restaurant“ mit der GPS-Koordinate des verwendeten Smartphones kombiniert, um de facto die Anfrage „italienisches Restaurant in Berlin“ zu generieren). Diese Möglichkeit haben Discovery-Systeme in der Regel nicht, da Ortsinformationen in den meisten Fällen keinerlei Bedeutung für eine Recherche haben dürften.

Dieses einführende Beispiel zeigt die Herausforderung für Suchmaschinen mit wenig Information seitens der Nutzenden umgehen zu müssen, aber trotzdem relevante Ergebnisse liefern zu müssen. Es ist deshalb wichtig, dass eine UI angeboten wird, welche die Nutzenden dabei unterstützt, ihren Informationsbedarf möglichst präzise zu formulieren, ohne selber wesentlich mehr Suchterme eingeben zu müssen. Hinzu kommt das Problem der Abdeckung. Während es ursprünglich, da das WWW wesentlich kleiner als heute war, genügte, möglichst relevante Webseiten zu präsentieren, müssen heute Bewertungen über den Inhalt der Seiten (z.B. deren Glaubhaftigkeit, Accessibility etc.) sowie eine Kontextualisierung der Suchanfrage in die Generierung der Suchanfrage mit einfließen. Diese Herausforderung lässt sich, wenn auch in einer kleineren Skalierung, auf den Bereich der Discovery-Systeme übertragen. Folgt man Blocks<sup>30</sup> Argumentation ist ein wesentliches Argument für Discovery-Systeme die Größe ihrer zugrundeliegenden Indizes, ergo: deren Abdeckung eines stetig wachsenden (virtuellen) Bestands der Bibliotheken.

<sup>28</sup> Hearst, M. A. (2009); Sec. 4.1.2f.

<sup>29</sup> Niu, X., Zhang, T., & Chen, H. (2014).

<sup>30</sup> Block, N. (2015).

Die Herausforderung besteht also folglich darin, aus einer knappen Suchanfrage, relevante Treffer, oder besser: Empfehlungen, abzuleiten, welche sich aus dem Kontext des Informationsbedürfnisses ergeben. Hierbei reicht es nicht aus, allein auf den Inhalt oder die Metadaten des nachgewiesenen Bestands zu setzen. Vielmehr müssen auch Strukturen bzw. Kontextinformationen mit einbezogen werden. Dass die Strukturierung von Inhalten nicht die Stärke von in Bibliotheken eingesetzten Suchmaschinen ist, wurde oben bereits dargelegt.

Es gilt also, sich in diesem Bereich besser aufzustellen, wobei die Startbedingungen hier für Bibliotheken besser sind als für die Suchmaschinen des WWW, wenn man sich auf seine Stärken, die Kurations- und Beratungskompetenz, besinnt. Das WWW besteht trotz Initiativen wie dem Semantic Web noch zum großen Teil aus unstrukturierten oder semi-strukturierten Daten, während bibliographische Daten dank Datenmodellen wie Functional Requirements for Bibliographic Records (FRBR) oder dem Regelwerk Resource Description and Access (RDA), zumindest konzeptionell, als Graph vorliegen.

Trotzdem hinderte dieser Umstand die WWW-Suchmaschinenbetreiber nicht daran, auf Grundlage der ungünstigen Datenausgangslage, komplexe Graphen zur Wissensrepräsentation („knowledge graphs“) zu modellieren, um neben der an ihre Grenzen stoßenden Volltextsuche weitere Daten heranziehen zu können, um Anfragen zu expandieren bzw. besser kontextualisieren zu können. Hierdurch wird ein Mehrwert für Nutzende generiert, ohne dass diese auf den Komfort einer kurzen Stichworteingabe verzichten müssen, wie Abbildung 5 mithilfe der Anfrage „john lydon“ illustriert.

Beide Suchmaschinen erkennen die Person „John Lydon“ inkl. ihres Pseudonyms „Johnny Rotten“, mit welcher sie als Sänger der Bands „Sex Pistols“ bzw.

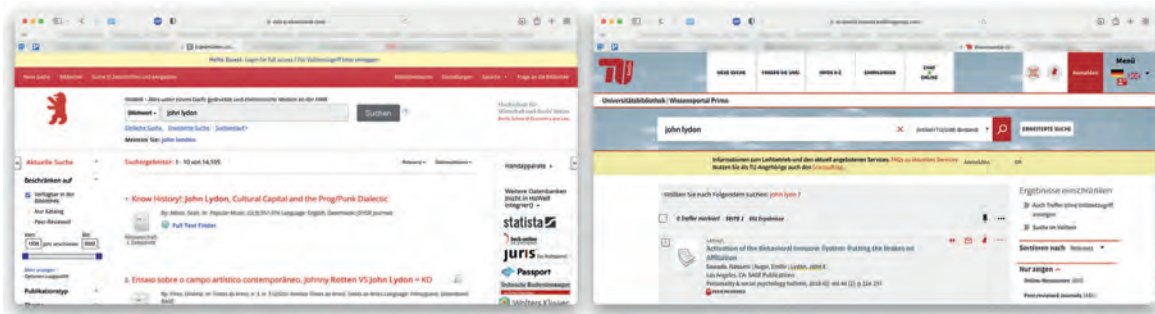


Abbildung 6:  
Ergebnislisten  
der HWR Berlin  
(EDS) und der TU  
Berlin (Primo) für  
die Anfrage „john  
lydon“ (v.l.n.r.)

„Public Image Ltd.“ tätig ist. Während Google (Abbildung 5, rechts) die traditionelle Listensemantik der Ergebnisliste nur zaghaft aufbricht, bindet Bing (Abbildung 5, links) Informationsblöcke über sein Werk, politische Einstellungen etc. direkt aus der Wikipedia ein, so dass Nutzerinnen und Nutzer keinen weiteren Klick ausführen müssen, um sich einen Überblick über die Person zu verschaffen. Dem Kontext angemessen werden weitere relevante „Fachdatenbanken“ wie YouTube, Spotify sowie diverse Social-Media-Netzwerke eingebunden. Dies ermöglicht eine direkte Exploration des Wirkens des Musikers, so dass bereits aus der ersten Trefferseite deutlich wird, dass es sich um einen bedeutenden Punk-Musiker handelt.

Abbildung 6 stellt diesen Ergebnissen das nüchterne Resultat zweier willkürlich gewählter Discovery-Systeme gegenüber. Die Systeme verharren in der klassischen Listendarstellung und nutzen offenbar keinerlei Kontextwissen aus, um Nutzerinnen und Nutzer Möglichkeiten zur Disambiguation der Suchterme „john lydon“ zu bieten. Während das System der HWR immerhin auf Literatur über John Lydon verweist, präsentiert das Wissensportal der TUB Werke verschiedener Autorinnen und Autoren mit dem Nachnamen „Lydon“. Als Disambiguation wird ein offenbar auf der Levenshtein-Distanz basierender Hinweis eingeblendet, der maximal Tippfehler abfängt. Diese Beispiele machen erneut deutlich, dass ausschließlich zeichenbasierte Inhalte verschiedener Felder im Index Berücksichtigung finden. Eine virtuelle Beratungsleistung findet nicht statt. Überträgt man das Bild der Bibliothek als Beratungsort ins virtuelle, so ist dieser Auftrag gescheitert. Die „Laien“ der WWW-Suchmaschinen sind hier erfolgreicher, da sie auf moderne Technik zurückgreifen können.

Das Festhalten am Query-Response-Paradigma und der Fixierung auf die einfache Listendarstellung ist sicherlich auch deren einfacher Nutzbarkeit zu verdanken. Proprietäre Anbieter müssen eine Vielzahl

von Kundenwünschen bedienen und können demnach nur eingeschränkte Anpassungsmöglichkeiten anbieten.

Diese Anforderung der kostensenkenden Generalisierung in einem sich immer mehr konsolidierenden Markt<sup>31</sup> steht jedoch der Wirklichkeit von Bibliotheken diametral gegenüber, die spezifische Inhalte kontextualisieren und fachspezifisch nachweisen müssen. So verfügt z.B. die Staatsbibliothek zu Berlin über eine Vielzahl an fachspezifischen Nachweissystemen, die materialspezifische Zugänge bieten und Nutzerinnen und Nutzer bestmöglich bei der Interaktion mit den Beständen unterstützen sollen. Zumeist finden sich diese Informationen nicht in den Discovery-Systemen wieder und wenn, dann unter Wegfall relevanter Informationen, da die Inhalte in das Feld-basierte Schema des Discovery-Systems gepresst werden müssen. Medien, wie z.B. Abbildungen, deren Relevanz durch Ähnlichkeit bestimmt wird, und sich damit nicht für die Darstellung in einer Liste eignen<sup>32</sup>, können allein aus technischen Gründen nicht in ein solches Discovery-System integriert werden.

Diese Beispiele machen deutlich, dass Bibliotheken Gefahr laufen, den Anschluss zu verlieren, wenn es ihnen nicht gelingt, das traditionelle Query-Response-Paradigma mitsamt der traditionellen Ergebnislistendarstellung zu überwinden. Die kommerziellen Discovery-Anbieter haben sich offenbar für das Verharren auf einer technischen Basis der frühen 2000er Jahre entschieden, die sich immer mehr von den Nutzererwartungen entfernen wird – gerade, wenn die zugrundeliegenden Indizes weiterwachsen. Dieses Phänomen ist nicht neu und konnte bereits beim Niedergang der OPACs beobachtet werden.

Versteht man eine Bibliothek als innovative Einrichtung, die Wissen auch durch Empfehlungen zugänglich macht, müssen folglich eigene „Boutique“-Recommender-Systeme mit den oben skizzierten Möglichkeiten entwickelt werden, um das volle Potential der eigenen Datenmodelle nutzen zu können. Dies

31 <https://americanlibrariesmagazine.org/2022/05/02/2022-library-systems-report/>

32 Santini, S. (2012). Because not all displays are lists. In Proceedings of the 2nd ACM International Conference on Multimedia Retrieval (S. 101-108). ACM.

schließt explizit nicht den Rückgriff auf kommerzielle Indizes aus. Allerdings muss deutlich werden, dass diese Indizes nur eine Datenquelle sind, aus der Informationsbedarfe gedeckt werden können. Bibliotheken müssen hier kuratierend tätig werden und die im Analogen existierende Expertise ins Virtuelle tragen, um Mehrwerte für Nutzerinnen und Nutzer zu schaffen, die domänenunspezifische Suchmaschinen nicht bieten.

Die dazu notwendigen Werkzeuge stehen in der Regel als Open Source bereit. Eine Named-Entity-Recognition ermöglicht es schon heute, den Kontext von Suchanfragen zu erkennen. Im obigen Beispiel ist es ein leichtes, „John Lydon“ als Person zu identifizieren und im Rahmen der Disambiguierung Nutzerinnen und Nutzer zu befragen, ob sie den Musiker, seine Verbindung zur Punk-Musik oder einen wissenschaftlichen Autor suchen. Die GND, Wikidata oder neue Initiativen wie OpenAlex<sup>33</sup> bieten hier entsprechende Anknüpfungspunkte, um Anfragen entsprechend zu expandieren.

Auch der datenschutzrechtlich<sup>34</sup> kompatible Rückgriff auf Anfragelogs kann durch Bibliotheken implementiert werden, um verwandte Suchanfragen zu identifizieren. Dies geht weit über die teilweise vorhandenen Ähnlichkeits-basierten Vorschlagsysteme hinaus, wie sie z.B. VuFind auf Basis einzelner Werke anbietet<sup>35</sup>.

Am Beispiel von VuFind wird deutlich, dass es nicht nötig ist, jede Funktion komplett neu zu erfinden, wenn man auf Open Source setzt. So konnte z.B. am 2017 eingestellten Sowiport gezeigt werden, dass sich VuFind erweitern lässt, um Query Expansions anhand von fachspezifischen Thesauri und wirkungsvolle Re-Ranking-Funktionen zu ermöglichen<sup>36</sup>. Ebenso bietet sich der Rückgriff auf Klassifizierungsdaten und Werkzeuge an, welche aus der Verwendung des maschinellen Lernens entstehen, wie z.B. aus dem Projekt „Erschließungsmaschine“ der DNB<sup>37</sup> oder Qurator<sup>38</sup>. Es ist nicht so, dass die Bibliotheken zu wenig Daten, Expertise oder Zugangsmöglichkeiten zu Nutzendenfeedback haben – diese Quellen müssen nur auch genutzt werden, um innovativ zu bleiben. Letztendlich haben sich Boutiquen auch ihren Platz durch Spezialisierung und Beratungskompetenz gesichert. |



**Prof. Dr.-Ing. David Zellhöfer**

Professor für Digitale Innovation in der öffentlichen Verwaltung  
HWR Berlin, Fachbereich 3  
david.zellhoefer@hwr-berlin.de

33 <https://docs.openalex.org>

34 Auch dieser Aspekt muss aufgrund des Platzes ausgeklammert werden, soll jedoch nicht, gerade mit Hinblick auf die in letzter Zeit bekannt gewordenen übergreifigen Tracking-Mechanismen der großen Verlagshäuser, unterschlagen werden.

35 <https://vufind.org/vufind/features.html>

36 Hienert, D., Sawitzki, F., & Mayr, P. (2015). Digital Library Research in Action: Supporting Information Retrieval in Sowiport. D Lib Mag., 21(3/4).

37 <https://blog.dnb.de/erschliessungsmaschine-gestartet/>

38 <https://github.com/qurator-spk>

NEUERSCHEINUNG



2022, Band 83  
Brosch., 264 Seiten  
978-3-9821824-6-9  
€ 29,50

Linda Freyberg | Sabine Wolf (Hrsg.)

## Smart wird man nur gemeinsam

Partizipative, agile und innovative Ansätze für die Bibliothek der Zukunft

Der Nachfolgebund zu dem 2019 erschienenen Band „Smart Libraries – Konzepte, Methoden und Strategien“ rückt die Partizipation stärker in den Vordergrund. Die Ereignisse der letzten zwei Jahre zeigen auf, dass die erfolgreiche Transformation in eine zukunftsfähige Bibliothek nur als gemeinsame Aufgabe gelingt, wenn das Wissen aller Beteiligten in den Prozess einfließen kann.

In diesem Band wird das Thema „Partizipation“ daher – auch kritisch – aus unterschiedlichen Blickwinkeln betrachtet.

Bestellungen unter [www.b.i.t.online.de](http://www.b.i.t.online.de)