

# Suchunterstützung in akademischen Suchmaschinen

Maria Zens<sup>1</sup>

Frank Sawitzki<sup>2</sup>

Philipp Mayr<sup>2</sup>

*GESIS – Leibniz-Institut für Sozialwissenschaften, Köln*

<sup>1</sup>*Fachinformation für Sozialwissenschaften*

<sup>2</sup>*Wissenstechnologien für Sozialwissenschaften*

*Unter Sachsenhausen 6-8, 50667 Köln*

*E-Mail: {maria.zens; frank.sawitzki; philipp.mayr}@gesis.org*

**Abstract.** Der Beitrag schließt an die Ausarbeitungen zu wissenschaftlichen Suchmaschinen, Query Understanding und Spezialsuchen der Bände 1 und 2 an. Es soll gezeigt werden, wie durch die Konvergenz von qualitativem Fach-Content und Suchtechnologien Mehrwerte gerade für Experten-Suchen generiert werden können. Die Beispiele aus unterschiedlichen akademischen Suchmaschinen (u.a. BASE, Web of Knowledge, Pubmed, Scopus, sowiport, Google Scholar und Deutsche Digitale Bibliothek) sollen das illustrieren, insofern sie grundsätzliche Fragen und Lösungsvorschläge zeigen, die aber über den einzelnen Anwendungsfall hinausweisen. Als in der Praxis erprobte State-of-the-Art-Dienste werden sie gleichwohl mit konkreten Beschreibungen der informationstechnischen Grundlagen untermauert.

**Keywords.** Wissenschaftliche Suchmaschinen, Suchstrategien, Browsing, Interaktives Retrieval, Monitoring, Relevanz, Query Understanding, Explorative Suche, Ergebnispräsentation, Universal Search, Usability

## Einleitung<sup>1</sup>

Am Anfang jeder Forschung stehen Suchen und Finden von Informationen. Das wissenschaftliche Interesse stellt dabei besondere Anforderungen an Präzision, Qualität, Aktualität, Originalität der Treffer und auch an die Möglichkeiten der nachhaltigen Nutzung und Weiterverarbeitung, denn im Gegensatz zu vielen alltäglichen Informationshandlungen, die meist konsumtiv bleiben, werden die Ergebnisse wissenschaftlicher Recherchen regelmäßig gespeichert, zitiert und gehen in die Produktion neuen Wissens ein. In diesem Kapitel geht es um die Unterstützung dieser Suchprozesse in Fachangeboten wie Portalen, Datenbanken und Digitalen Bibliotheken und insbesondere darum, wo Retrieval-Technologien und Trefferaufbereitung über das hinausgehen, was allgemeine Suchen erfordern und allgemeine Suchmaschinen leisten.

---

<sup>1</sup> Die Abschnitte verteilen sich wie folgt: Einleitung, Kap. 1-3 sowie Zusammenfassung (M. Zens); Kap. 4-7 (F. Sawitzki/ P. Mayr).

Der Beitrag schließt damit an die Ausarbeitungen zu wissenschaftlichen Suchmaschinen [1] [2], Query Understanding [3] und Usability [4] sowie Spezialsuchmaschinen [5] an, wie sie in den beiden ersten Bänden des Handbuchs zu finden sind. Während diese Beiträge vor allem die Aspekte wissenschaftlicher Content [2] und Basistechnologien [1] behandelt haben, möchten wir den Bereich des interaktiven Informationshandelns in den Fokus nehmen – also die situative Sicht des Nutzers einnehmen und fragen, wie typische Suchen optimal unterstützt werden können. Diese Perspektive steht im Einklang mit einer stärkeren Berücksichtigung von sozialen und prozessualen Grundlagen in der Informationswissenschaft, wie sie Ingwersen und Järvelin vorgeschlagen haben [6], und zielt auf den Bereich zwischen Akteuren und Technologie, „the interaction between people and content in information systems“ [7].

Wissenschaftliche Suchmaschinen und Fachdatenbanken leisten für den Bereich akademischer Recherchen Komplementäres; wir möchten spezifische Dienste zur Suchunterstützung beleuchten, mit denen diese Leistungen aggregiert werden können. Es wird gezeigt, wie durch die Konvergenz von Fach-Content und Suchtechnologien Mehrwerte gerade für „Experten-Suchen“ generiert werden können. Wissenschaftliche Suchen werden hier als solche Experten-Suchen verstanden, die nicht nur besondere Anforderungen an die Qualität und Aktualität der Inhalte stellen, sondern darüber hinaus den spezifischen Regeln des jeweiligen disziplinären Feldes folgen. Was für die zeitweise heftig geführte Diskussion über die Evaluation von Forschungsausgaben gilt – die einzelnen Fächer folgen ihren eigenen Logiken und Traditionen und müssen danach beurteilt werden –, gilt auch für die Informationssuche und Wissensorganisation. Das wird unmittelbar einleuchtend, wenn man sich die unterschiedlichen Anforderungen an Retrieval und Präsentation von Informationsangeboten in beispielsweise der Kunstgeschichte oder der Chemie vor Augen führt.

Die in unserem Beitrag angeführten Beispiele, die auf verschiedene wissenschaftliche Informationsangebote – u.a. das sozialwissenschaftliche Fachportal sowiport – verweisen, sollen die Spezifik und auch eine gewisse Bandbreite der wissenschaftlichen Suchmöglichkeiten illustrieren und dabei grundsätzliche Fragen und Lösungsvorschläge zeigen, die aber über den einzelnen Anwendungsfall hinausweisen (s. Tabelle 2). Als in der Praxis erprobte State-of-the-Art-Dienste werden sie gleichwohl mit Beschreibungen der informationstechnischen Grundlagen untermauert.

Im ersten Teil unseres Beitrags sollen einige Spezifika und grundsätzliche Tendenzen wissenschaftlicher Informationssuche umrissen werden (Abschnitt 1), gefolgt von einer Typologie zentraler Suchstrategien (Abschnitt 2) und einem Abschnitt, der das Identifizieren und Auswählen von relevanten Informationen durch Relevanzzuweisung, Suchpräzisierung und strukturierte Trefferpräsentation thematisiert (Abschnitt 3). Der zweite Teil des Beitrags greift diese Struktur auf und soll exemplarisch darstellen, wie Anforderungen der fachwissenschaftlichen Logik in Retrievalangebote und Suchinfrastrukturen umgesetzt werden können. Das geschieht auf Ebene des Suchinterfaces, das ‚Standardsituationen‘ der Recherche und fachspezifische Suchen unterstützen muss (Abschnitt 5), auf Ebene der Trefferpräsentation (Abschnitt 6) und des Rankings sowie hinsichtlich der „intelligenten“ – im Sinne von fachwissenschaftlich relevanten – Expansion individueller Suchstrategien, wie sie beispielsweise von der Dokumentenebene (Abschnitt 7) ausgehen kann.

## 1. Wissenschaftliches Recherchieren als Experten-Suche

Das Recherchieren wissenschaftlicher Informationen ist zentraler Bestandteil akademischer Arbeit und steht am Anfang des Forschungszyklus. Als erstes denkt man hier natürlich an die Literaturrecherche, die auch sicher im Mittelpunkt steht, hinsichtlich der zur Verfügung stehenden Informationsinfrastrukturen am besten ausgebaut und auch am besten erforscht ist. Es wäre aber zu kurz gegriffen, das Korpus auf Forschungsliteratur und ihre Nachweise zu reduzieren, denn auch andere Informationstypen (Projekte, Tagungen, Quellenkorpora etc.), darunter auch solche, die nicht unmittelbar textuell repräsentiert sind (Datenreihen, Graphiken, Artefakte und ihre digitalen Abbildungen, Video- oder Audiodateien) werden benötigt. Unabhängig von Informations- oder Dateityp sind alle auf diese Einheiten zielenden wissenschaftlichen Suchen Experten-Suchen, insofern sie Spezialwissen erfordern und abfragen. Eine erfolgreiche Recherche im akademischen Bereich erfordert Informationskompetenz, die nicht nur den Umgang mit Tools, sondern auch aktives Domänenwissen umfasst: Fachvokabular, Standards der disziplinären Wissensorganisation usw.

Wie finden erfahrene Nutzer die für sie wichtigen Informationen in Datenbanken, Portalen und Bibliotheken? Wissenschaftler arbeiten in einem hybriden Informationsraum, in dem die Wertschätzung für die klassischen Standards dokumentarischen Arbeitens und bibliothekarischer Services und die selbstverständliche Nutzung der Suchmaschinenwelt ineinandergreifen.

Wissenschaftliches Suchen beginnt also nicht mit der Erfindung von Internet-Suchmaschinen, was man angesichts der Allgegenwart derselben leicht vergisst. Eine neuere Publikation, die sich in historischer Perspektive mit Systemen wissenschaftlicher Recherche beschäftigt, hat beides – die Tatsache und das Vergessen – in ihrem Titel auf den Punkt gebracht: „Vor Google. Eine Mediengeschichte der Suchmaschine im analogen Zeitalter“ [8]. Zweifellos erfahren aber auch die Techniken und Verfahren wissenschaftlicher Informationssuche einen tiefgreifenden Wandel. Ging es zunächst darum, bekannte und bewährte Informationsstrukturen (z.B. Bibliothekskataloge) auf digitale Angebote umzustellen, die die gleichen Recherchen mit anderen Mitteln ermöglichen sollten, haben Retrievaltechnologien die Systeme wissenschaftlicher Recherche und den Informationshabitus beeinflusst. Gerade in den letzten Jahren sind einige empirische Studien vorgelegt worden [9] [10] [11] [12] [13] [14] [15], die sich genauer mit Informationssuchen im wissenschaftlichen Bereich und den Veränderungsprozessen, denen diese unterworfen sind, auseinandergesetzt haben. Ohne auf die vielfältigen Ergebnisse im Einzelnen eingehen zu können, seien einige der zentralen Punkte herausgegriffen, die wir auch in eigenen Untersuchungen bestätigt gefunden haben. GESIS hat beispielsweise in leitfadengestützten Tiefeninterviews erfahrene Sozialwissenschaftler (Professoren und wissenschaftliche Mitarbeiter/ Post-Docs) zu ihrem Informationsverhalten, ihren individuellen Recherchepfaden für Forschung und Lehre sowie nach Desideraten in der Informationsversorgung befragt.

(1) Die digitale Informationssuche und die digitale Verfügbarkeit der Ressourcen werden immer wichtiger. Das ist insofern nicht trivial, als die Print-Publikation weiterhin einen besonderen Stellenwert genießt und – zumindest in den Geistes- und Sozialwissenschaften – die Beschränkung auf einen rein digitalen Informationszyklus derzeit nicht denkbar erscheint. Digitale Veröffentlichungs- und Rezeptionsformen lösen analoge also nicht einfach ab.

(2) Suchmaschinen werden als allgemeine Suchoberfläche auch für wissenschaftliche Inhalte genutzt, und, vielleicht noch entscheidender, das Suchverhalten orientiert sich an diesen Angeboten. Für akademische Informationsanbieter heißt das zum einen, ihre Inhalte zumindest teilweise über die Suchmaschinen verfügbar zu machen (hierzu auch: [2]), zum anderen, ihre eigenen Such-Interfaces an die Standards anzugleichen, die vor allem Google setzt. Die Anforderungen, die auch im wissenschaftlichen Kontext an die Technik gestellt werden, beinhalten einen einfachen, intuitiv verständlichen Zugriff, Schnelligkeit und Annehmlichkeit bei gleichzeitigem Anspruch an einen qualitativ hochwertigen Inhalt und eine hohe Präzision und Spezifik der Ergebnisse.

(3) Ein weiterer wesentlicher Aspekt für die Einschätzung von Experten-Suchen ist, dass die gezielte Nutzung von „quality resources“ [11] mit der Berufserfahrung zunimmt (vgl. auch [10]), mit anderen Worten: fachspezifische Quellen- und Informationskompetenz sind Bestandteil des professionellen Wissens. Das ist zu berücksichtigen, wenn ein Informationsangebot zugleich Studienanfänger und erfahrene Wissenschaftler ansprechen soll.

(4) Auch wenn die Bedeutung von Bibliotheken bei der Informationssuche stark zurückgegangen ist, gilt das nicht für die Beschaffung (Lizenzierung von Datenbanken, Ausleihe). Die Verknüpfung mit Verfügbarkeitsdaten und Rechtemanagement ist also wichtig.

(5) Die genutzten Informationsmittel werden der wissenschaftlichen Suchstrategie subsumiert, nicht umgekehrt. Das bedeutet: Wissenschaftler interessieren sich für inhaltlich relevante Informationen, aber i.d.R. nicht für Suchmasken oder Thesauri; sie schätzen an Retrieval-Instrumenten den hohen Nutzen, den sie bringen, und den geringen Aufwand, den sie erfordern. Die unterschiedlichen Einschätzungen von Nutzern und Anbietern, was gute Informationsangebote ausmacht, zeigt anschaulich die Untersuchung des OCLC „Online Catalogs: What Users and Librarians Want“ [13].

(6) Die vielleicht wichtigste Erkenntnis ist die große Bandbreite des Informationsverhaltens, in der Studie des JISC heißt es zusammenfassend: „one size doesn't fit all“ [12]. Zu dieser Differenziertheit gehört auch, dass die disziplinären Eigenheiten des Informationsverhaltens wichtig sind (s. auch [10]) und von Experten explizit als solche formuliert werden („Fachkultur“).

Gerade in diesem letzten Befund verbirgt sich die Herausforderung für wissenschaftliche Informationsanbieter, denn wie soll man möglichst spezifisches Recherchieren für möglichst viele individuelle und anspruchsvolle Nutzer anbieten? In der GESIS-Interviewreihe mit erfahrenen Sozialwissenschaftlern wurde neben anderem eine große Differenziertheit der Suchstrategien deutlich – natürlich suchen alle Befragten nach Literatur, aber wie und wo sie das tun, wie sie nach Qualität und nach Relevanz bewerten, ist doch recht unterschiedlich. Das Informationsverhalten erwies sich als vielfältig hinsichtlich der Inhalte und bevorzugten Dokumenttypen, es gab eine breite Streuung der genutzten Informations-Tools und Quellen und sehr differenzierte Suchstrategien. Es wurde auch nach Unterschieden der Suche für verschiedene Zwecke (Forschung oder Lehre, neues Forschungsfeld oder bekanntes Thema) gefragt und das Ergebnis war, dass insbesondere „exploratives“, forschendes Suchen sehr individuellen Pfaden folgt. Insbesondere diese Aspekte von Ranking, Relevanzsetzung und explorativer Suche stellen erhöhte Anforderungen an die Suchtechnologie und die Präsentation der Ergebnisse in wissenschaftsspezifischen Angeboten. Allgemeine Suchmaschinen hingegen werden – weil sie effizient und schnell sind – auch im

wissenschaftlichen Bereich für gezielte Suchen („lookup“, „known-item-search“, siehe 2.2) genutzt.

Welche Konsequenzen hat das für die Aufbereitung von Informationen für Experten? Eine Leithypothese ist, dass fachspezifisches Wissen durch Verknüpfungslogiken, die den jeweiligen Anforderungen des disziplinären Feldes folgen, adäquater präsentiert und präziser gefunden werden kann, als es z. B. allgemeine Suchmaschinen ermöglichen. Während bei Alltagssuchen meist die Präzision, mit der schnelle und eindeutige Suchergebnisse erzielt werden, im Vordergrund steht, ist für wissenschaftliche Suchen auch das explorative Erschließen neuer Forschungskontexte wichtig. Diese – die eigentlich „forschende“ – Suchstrategie geht Umwege, basiert auf trial-and-error, ist komplex und hat ganz unterschiedliche strukturelle und inhaltliche Ausgangspunkte („seeds“). Gerade die individuellen und neuen Forschungsinteressen lassen sich weniger gut mit eindeutigen Antworten bedienen oder Rankings, die auf Häufigkeiten beruhen. Für Experten sind genaue und tiefe Informationen wichtig, die eine individuelle Relevanzzuweisung ermöglichen, und es ist wichtig, vielfältige Ergebniskontexte mit Anschlussinformationen zu erhalten, die möglichst nahtlos zur Verfügung stehen, also keine neue Suche erfordern.

## **2. Suchstrategien identifizieren und „verstehen“**

Optimale Suchunterstützung hat zum Ziel, die Varianz von Suchanfragen möglichst gezielt zu bedienen. Die primäre Zielgruppe akademischer Suchangebote sind Fachwissenschaftler. Wissenschaftler sind Suchexperten, insofern Recherchen integraler und notwendiger Bestandteil wissenschaftlichen Arbeitens sind und eine hohe fachspezifische Informationskompetenz unterstellt werden darf, was nicht heißt, dass Wissenschaftler immer in der Lage oder überhaupt willens wären, die technischen Retrievalmöglichkeiten von Portalen und Suchmaschinen optimal zu nutzen. Das macht Suchunterstützung einerseits schwierig – der Anspruch ist hoch und hochspezifisch – eröffnet aber auch die Möglichkeit, das Wissen der Nutzer durch interaktive Angebote einzubinden und Ergebnisse so iterativ zu verbessern. Experten sind in dieser Hinsicht im Vorteil. Der gezielte Einsatz präziser fachterminologischer Suchwörter und das Erkennen von Anschlussnavigationen, die von einem Suchergebnis ausgehend explorieren, ohne einer neuen Initialsuche zu bedürfen, können systemseitig unterstützt werden.

Zunächst einmal ist es jedoch wichtig, die Struktur wissenschaftlicher Suchanfragen zu „verstehen“, Standardfragen und -aktivitäten zu identifizieren, diese in Retrievalstrukturen umzusetzen, um sie dann mit weiterführenden Browsingmöglichkeiten und Suchpfaden zu verknüpfen.

### *2.1. „Search tasks“, Suchziele und Suchaktionen*

Wissenschaftliche Suchen sind zwar in einen anderen Kontext der Informationssuche und Wissensproduktion eingebettet als Alltagssuchen, die zugrundeliegenden „search tasks“ lassen sich aber strukturell ähnlich beschreiben. Das Suchinteresse wird vom Informationsbedarf bestimmt, der nach Frants et al. ([16], ins Deutsche übertragen bei Stock [17]) konkret und auf die Beantwortung einer präzise zu

stellenden Frage gerichtet oder eher problem- und aufgabenorientiert und damit offener und „explorativer“ sein kann. Eine generelle Typologie von Anfrageformen im Web bei Broder [18] unterscheidet Aktionen, die als „transactional“, „informational“, „navigational“ klassifiziert werden. Für Fragen der wissenschaftlichen Suchunterstützung sind vor allem die mit Informationsgewinnung verbundenen Aufgaben („informational tasks“) relevant, insofern sich hier das eigentlich forschende Interesse artikuliert, das auf neues oder neukombiniertes Wissen zielt. Die grundlegende Unterscheidung von explorierenden Suchen („exploratory search“) und dem Aufrufen oder „Nachschlagen“ bekannter Informationen („lookup“), wie sie Marchionini [19] vornimmt, könnte man als weitere Differenzierung der Kategorie von „informational search tasks“ verstehen. Marchionini selbst rechnet der explorativen Suche die Aktionen „learn“ und „investigate“ zu, die er als einander in der Praxis überlappend ansieht. Die bereits aufgeführte Project Information Literacy-Studie [9], die sich mit dem Suchverhalten von Studierenden beschäftigt, entwirft ein viergeteiltes Schema mit den Suchtypen „big picture“ (Hintergrundinformationen), „language“ (Begriffe und Terminologie), „situational“ (etwa: Reichweite der Suche) und „information gathering“ (Auswählen und Beschaffen relevanter Information). Im Folgenden sollen einige Suchoperationen in ihrer für die wissenschaftliche Suche typischen Ausprägung skizziert werden.

## 2.2. „Lookup“ und „known-item-search“

Mit „known-item-search“ wird eine Suchoperation bezeichnet, die das (Wieder-)Finden von Bekanntem zum Ziel hat. Hier ist eine möglichst große Präzision gefragt.

Wie kann gezielt nach Bekanntem gesucht werden? Mit möglichst vielen Informationen wie einer kompletten bibliographischen Angabe – was die Gefahr eines „null Treffer“-Ergebnisses birgt – oder mit wenigen, aber möglichst spezifischen Angaben, was die Frage aufwirft, welche das sinnvollerweise sind. Was ist der bekannte Teil eines „known items“ und was ist ein hinreichend präziser Teil, um eine gezielte Suche erfolgreich abzusetzen? In der Regel werden (Autor-)Namen, Titelstichwort, Erscheinungsjahr, Projektakronym o.ä. Teil einer solchen Query sein, also die Elemente, die traditionell zum Nachweis verwendet und in Datenbanken über verschiedene Felder gezielt adressierbar sind.

Ein Beispiel soll das verdeutlichen: Wenn Habermas' Standardwerk zur diskursiven Formation moderner Gesellschaft (Jürgen Habermas: Strukturwandel der Öffentlichkeit. Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft. Neuwied: Luchterhand 1962) gesucht wird, sind „Habermas 1962“ oder „Habermas Strukturwandel“ zielführende Queries. Der von Habermas geprägte und in seiner Nachfolge fest verankerte Begriff „Strukturwandel“ ist präziser als der der „Öffentlichkeit“ und wird als Stichwort bevorzugt werden. Das sind zugleich – was nicht verwundern dürfte – häufige Formen der Kurzzitation, die ja ebenso die Funktion haben, das zitierte Werk möglichst kurz und genau zu bezeichnen. Die ‚Verstehensleistung‘ bestünde beispielsweise darin zu erkennen, dass es bei einer Suche nach „Habermas 1962“ – obwohl der Term aus Personennamen und Jahreszahl besteht – um ein Drittes geht, nämlich das Auffinden einer konkreten Publikation, die durch diese Kombination gut eingegrenzt wird. Für einen an den Umgang mit Fußnoten und Zitaten gewöhnten Forscher bedarf das aufgrund der eingeübten Konvention keines Nachdenkens, aus der Zeichenkette selbst ist es nicht unmittelbar zu ermitteln.

Ein solches „Verstehen“ ist aber wichtig für die Suchunterstützung, wenn z.B. weiterführende Informationen und Empfehlungen angeboten werden sollen. In unserem Beispiel wären Arbeiten, die sich auf das Buch „Strukturwandel der Öffentlichkeit“ beziehen, relevanter als weitere Veröffentlichungen des Autors Habermas oder gar weitere Publikationen mit dem Erscheinungsjahr 1962. An dieser Stelle kann auch verdeutlicht werden, wie Domänenwissen die Query-Formulierung prägt [20] [21]. Ein Experte wählt die ‚richtigen‘ Suchbegriffe und er weiß umgekehrt auch, wie er nicht suchen sollte: Ein Kant-Kenner würde kaum jemals mit dem Term „Kant Kritik“ suchen – nicht nach einer der Kritiken Kants und erst recht nicht nach einer kritischen Abhandlung *über* den Philosophen. In jedem Fall würde er mit unpräzisen Treffermengen rechnen. Erfolgversprechender wären „Kant Urteilskraft“ (bzw. „...reine Vernunft“ / „praktische Vernunft“) für das erste Suchinteresse und „Kant Rezeption“ für das zweite.

Eine Variante des „lookup“ könnte man als „one-item-search“ bezeichnen, wenn es darum geht, Erstbestes oder Hinreichendes zu finden, das eine genau definierte und begrenzte Funktion erfüllt und den jeweiligen Qualitätskriterien genügt, ohne dass aber ein exakter Treffer bereits bekannt wäre oder es überhaupt ein bestimmter Treffer sein müsste. Gesucht werden könnten z.B. (irgend-)eine Publikation zur Lyrik und Epik des Cinquecento, ein Diagramm, das die demographische Entwicklung in der EU zeigt, oder ein Portrait Isaac Newtons, die als Beispiele oder Beiwerk genutzt werden sollen – dafür, dass Italien im 16. Jahrhundert eine avancierte Dichtkunst aufzuweisen hat, als Begründung für steigende Gesundheitsausgaben, als Illustration eines Vortrags über Große Männer in der Geschichte der Naturwissenschaften. Das Dokument selbst ist also nicht Anknüpfungspunkt für weitere Forschung. Das unterscheidet sich von einer wissenschaftlichen Beschäftigung mit Themen, für die dieselben Informationen auch eine Rolle spielen könnten, die aber den Stand der Forschung abbilden müssten und höhere Ansprüche an Qualität, Aktualität und Recall stellen würden, mit „Erstbestem“ also nicht zufrieden sein könnten: Das könnten für das genannte Beispiel eine Neuedition und historische Einordnung des „Orlando furioso“ sein, eine Evaluation demographischer Modellrechnungen, eine ikonologische Studie zur zeitgenössischen Repräsentation der Naturwissenschaften. Dieselben Elemente sind in dem einen Fall hinreichend und werden nicht weiter verarbeitet, im zweiten aber Teil des zu erforschenden Gegenstandsbereichs und mithin Ausgangspunkt für die weitere wissenschaftliche Beschäftigung. Eine solche wird, insbesondere, wenn es sich um ein neues Feld handelt, eher von explorativen Suchen (2.3) bestimmt.

„Lookup“-Aktionen scheinen auf den ersten Blick leichter zu bedienen zu sein als komplexere Fragestellungen, in der Praxis werden für die Nutzer hier aber allgemeine Suchmaschinen häufig die erste Wahl sein, weil sie schnell und effizient sind und oft hinreichend gute Ergebnisse liefern. Allgemeine Suchmaschinen sind jedoch weniger geeignet für explorative Suchen [22], die von wissenschaftlichen Suchmaschinen bzw. Fachangeboten besser unterstützt werden.

### 2.3. Explorative Suche

Konzepte von „explorativer Suche“ finden sich in zahlreichen Texten, explizit so bezeichnet z.B. bei Marchionini [19], titelgebend beim ACM SIGIR Workshop „Evaluating Exploratory Search Systems“ 2006 [23] sowie für ein Special Issue zu „Supporting Exploratory Search“ [24] – und auch bereits in den Texten von Bates [25]. „Explorieren“ – Suchen ohne präzise Vorstellung, was gefunden werden wird, aber mit

Kriterien, die Relevantes von Irrelevantem scheiden – ist praktisch ein Synonym für das wissenschaftliche Arbeiten. Als klassische „Schneeballsuche“ wird sie in der Einführungsliteratur jeder Studierendengeneration mitgegeben (stellvertretend für viele: [26][27][28]). Die Metaphern des „chaining“ und „berrypicking“ bei Bates [25] oder der von Merton populär gemachte Begriff der „serendipity“ [29][30] bezeichnen alle mehr (Bates) oder weniger (Merton) linear gedachte Suchmodelle, in denen Informationen in einem Feedback-loop aufgegriffen und zur Präzisierung, Veränderung oder Erweiterung des ursprünglichen Suchinteresses beitragen. Damit stellen sie sich außerhalb des einfachen Frage-Antwort-Paradigmas.

Der Begriff der explorativen Suche wird nicht immer eindeutig verwendet und gelegentlich sogar negativ konnotiert mit einem Mangel an Informationskompetenz oder Fachwissen in Zusammenhang gebracht, etwa im Sinne eines ‚hilflosen Herumirrens im Informationsdschungel‘. Reynolds et al. modellieren ihre Suchunterstützung an „non-expert users“ [31]. Kules und Capra betonen „uncertainty, ambiguity and discovery“ [32] als Kennzeichen. Die Besonderheit explorativer Suchen im wissenschaftlichen Kontext ist jedoch, dass sie zwar mit einem Defizit an Wissen zu tun haben, insofern häufig Forschungslücken der Ausgangspunkt sind, sie aber keineswegs einem Mangel an Informations- oder Domänenkompetenz entspringen, sondern eher dem Gegenteil. Es ist davon auszugehen, dass „explorative Suche in Informationskollektionen der Arbeitsweise von Wissenschaftlern und Technikern besonders entspricht und dass aktive Navigation im virtuellen Informationsraum eine sehr wirksame Suchtechnik zur Unterstützung dieses Vorgehens ist“ [33]. Solche Suchen sind immer dann besonders wichtig, wenn es darum geht, neue Themen zu bestimmen und zu erschließen. Sie sind gekennzeichnet durch iterative Recherche, sie sind komplex, nicht-linear, „unpräzise“ in der Query-Formulierung, sie können – strukturell – ganz unterschiedliche Ausgangspunkte haben und sehr individuelle Fragen stellen; sie sind im besten Falle neu und expandieren das eigene Wissensfeld.

Ein wichtiges Merkmal ist die Flexibilität der Suchaktion, die typischerweise zwischen Vertiefung und Erweiterung der Anfrage wechselt und dabei die Angebote des oder der Informationstools interaktiv nutzt. Das langsame Verfertigen der Strategie während des Suchens impliziert sog. „query drifts“ [34], Bewegungen, mit denen die Suchrichtung geändert wird, wenn ein relevanter Anknüpfungspunkt erscheint. Unterstützt werden explorative Suchen, indem vielfältige Ergebniskontexte („Verzweigungen“) angeboten werden. Das geschieht zwar auch über die Facettierung von Suchergebnissen (Verengung), wichtiger im Sinne des Iterations-Prinzips sind jedoch die erweiternden Pfade, die in digitalen Informationsangeboten idealerweise „klickbar“ sind. Beispiele hierfür sind die Referenzen einer Publikation, umgekehrt die Informationen zu Zitationen oder die Expansion mit Hilfe von Autornamen oder Sacherschließung. Neben diesen, von den Informationseinheiten selbst, der intellektuellen Erschließung in Datenbankangeboten oder durch Tagging erzeugten Verlinkungen, sind hier auch Empfehlungen von „ähnlichen Suchen“ (s. auch 6.4) und „ähnlichen Informationen/ Dokumenten“ (Punkt 7) zu nennen.

Wie solche Pfade konkret aussehen können, soll ein Beispiel zeigen: Wird mit dem Term „Gesundheitliche Ungleichheit“ gesucht, könnten – neben Forschungspublikationen zu diesem Thema – vielfältige Informationen von Interesse sein:

- Indikatoren und Datenreihen,
- internationale Studien,



- zentrale Autoren,
- zentrale Publikationsorgane („Kernzeitschriften“).

Naheliegend ist auch die auf übergeordnete Themen zielende Expansion:

- Theorien der Ungleichheit,
- normative Fragen nach Gerechtigkeit,
- Public Health,
- Prävention,
- medizinische Versorgung,

... oder Tabellen zu „healthy life years“ in verschiedenen Ländern oder Studien zu einzelnen Krankheitsbildern, zu Lebensstilen oder zur Bedeutung von Gesundheitsforschung in der Entwicklungspolitik – diese Liste ließe sich sicher noch verlängern und macht deutlich, dass Rankingmechanismen sich kaum eignen, um solch differenzierte Anfragekontexte zu bedienen. Auch Suchinterfaces werden unübersichtlich und letztlich unbrauchbar, wenn sie mit zu vielen Auswahloptionen überladen werden. In der Praxis scheint es daher sinnvoll, zum einem häufig genutzte Pfade für explorative Suchoperationen anzubieten (s. auch 3.1), zum anderen den Nutzern Optionen einzuräumen, solche Richtungsänderungen selbstständig einzuschlagen. Das stellt gewisse Anforderungen an die Trefferpräsentation, die zugleich übersichtlich und so strukturiert sein soll, dass Knoten und Abzweigungen ins Auge springen. Auch sollte die nahtlose Kombination von Such- und Browsingaktionen gewährleistet sein.

Möglichkeiten der thematischen Verfeinerung oder Erweiterung mit Suchtechnologien sind Facetten, Durchsuchen der Trefferliste, Termclouds oder das Editieren der Ausgangsquery. Browsingoptionen können sich an Referenzen und Zitation, Namen, Schlagwörtern/ Klassifikationen oder zugehörige sowie ähnliche Informationen anschließen. Das alles klingt vertraut und entspricht in vielem den Techniken, mit denen auch in der analogen Ära gesucht und gefunden wurde. Der große Vorzug digitaler Informationssuche liegt darin, all diese Operationen integriert anbieten und auf große Datenmengen anwenden zu können.

Die Kombination von Informationsbreite und -tiefe wird am deutlichsten, wenn auf Metadaten und Erschließung basierendes Retrieval mit der Verfügbarmachung von Volltexten verbunden wird. Ob ein Buch oder Aufsatz für die eigene Fragestellung relevant ist, lässt sich – zumindest für erfahrene Wissenschaftler – am ehesten mit einem Blick in den Text beantworten. Die Volltextsuche ist daher sowohl für die Ermittlung von Treffern als auch im Anschluss daran, als Scannen eines relevanten Textes hilfreich. Die Einbindung von externen Ressourcen wie Google Scholar, Google Books oder Verlagsangeboten mit möglichst weitgehenden Informationen ermöglicht, relevante Treffer weiter zu verfolgen.

Erfolgreiches „Explorieren“ beinhaltet auch Irrwege; deshalb sind alle Instrumente hilfreich, die Suchverläufe aufzeichnen oder ein komfortables „Zurück“ an verschiedene Stellen der Suchhistorie ermöglichen (wie Breadcrumbs, Aufrufen und Editieren von Suchen, s. auch Punkt 5.2). Überhaupt legt die „explorative Suche“ den Schwerpunkt auf das, was zwischen Frage und Antwort steht: „the path is the destination“ [35]. Solche Pfade durch Verlinkung heterogener Datenbestände anzubieten und durch entsprechendes Interface-Design [36] abfragbar zu machen, unterstützt Explorieren, indem das Wissen der Nutzer eingebunden wird.

## 2.4. Monitoring

Insbesondere für erfahrene Wissenschaftler ist das Beobachten der Entwicklung eines bekannten („ihres“) Themenfelds wichtig. Dazu gehört auch, die Rezeption der eigenen Arbeiten zu verfolgen, wie es beispielsweise mit Google Scholar möglich ist. Mail-Alerts und RSS-Feeds gehören heute beinahe zu den Standardausrüstungen – siehe die Aufstellung bei Plieninger [37]. Die automatisierten Angebote ergänzen die älteren redaktionell betreuten Formen wie Newsletter und Mailinglisten. Auch bei Feeds und Alerts ist die wichtigste Aufgabe, relevante Informationen herauszuschälen und Ballast auszugrenzen. Die Möglichkeiten, selbstkonfigurierte komplexe Suchen als Benachrichtigungsdienste abzuspeichern, ist hier sicher eine avancierte Form, die Spezialisten entgegenkommt.

Neben diesen personalisierten Angeboten, die mehr oder weniger feingranular gestaltet werden können, stehen eine ganze Palette „voreingestellter“ Monitoring-Möglichkeiten zur Verfügung. Zu nennen sind hier z.B. TOC-Alerts („table-of-content“-Alerts) – Inhaltsverzeichnisse, oft mit den Abstracts neuerschienener Zeitschriftenhefte, die von einzelnen Verlagen oder auch aggregiert abonniert werden können. Fachdatenbanken und Kataloge bieten Neuerwerbungs- bzw. Neuzuganglisten an, die über Klassifikationen gefiltert oder mit Stichwörtern durchsucht werden können. Die „digitale Auslage“ von Neuerscheinungen intelligent zu filtern und zusammenzustellen, ist letztlich nicht weniger aufwendig, als sich in der Welt physischer Informationsangebote zurechtzufinden.

Der Rückgriff auf thematische Informationsangebote, die Neuigkeiten aus verschiedenen Quellen aggregiert zugänglich machen – ein Beispiel hierfür sind Themen-Feeds, wie sie im Portal sowiport angeboten werden –, ist eine Alternative. Da solche Angebote aber einen Scope aufweisen, der notwendigerweise allgemeiner ist als das Interesse einer individuellen Fragestellung, wird eine anschließende Selektionsarbeit immer nötig sein.

Auch Suchen über eingeschränkte Zeiträume können in gewisser Weise Monitoring-Funktion erfüllen, müssen sich aber i.d.R. auf das bei großen Treffermengen zu unscharfe Suchkriterium des Erscheinungsjahrs beschränken. Genauer sind Suchmöglichkeiten, die Abfragen über das (Zugangs-)Datum ermöglichen oder Sortierungsoptionen nach Zugangs- respektive Erscheinungsdatum. Je nachdem, wie präzise die Metadaten diesbezüglich sind, erhält man gute Ergebnisse. Voraussetzung für solche Services auf Seiten der Informationsanbieter ist es, Aktualitätsabfragen mit möglichst genauen Zeitangaben, die über Suchmasken, Sortierung oder Facetten angesteuert werden können, zu unterstützen.

## 3. Finden unterstützen

Ziel einer jeden Suche ist das Finden. Eine Suche gilt erst dann als erfolgreich beendet, wenn ich das bekomme, was ich brauche. Das mag genau das sein, was ich gesucht habe, oder vielleicht noch besser: ich entdecke etwas, von dem ich gar nichts wusste. Das gilt nicht nur im wissenschaftlichen Kontext, aber hier spielen die stetige Erweiterung des Wissens und Neuverknüpfung von Informationen eine besondere Rolle.

### 3.1. Relevanzzuweisung zwischen ‚Standardsituationen‘ und Spezialfragen

„Finden“ ist also gleichbedeutend mit dem Erhalt eines relevanten und präzisen Ergebnisses. Wie gezeigt, können die Intentionen von Suchoperationen aber ganz unterschiedlich sein. Für Relevanz ist es entscheidend, *Suchanfragen* zu ermöglichen, die solche Intentionen bereits transportieren, *Daten* anzuzeigen, die für Relevanzzuweisung wichtige Informationen beinhalten, und die *Trefferpräsentation* so zu gestalten, dass nachträgliche Verfeinerungen angeboten werden. Saracevic [38] und Hjørland [39] haben die Dimensionen des Relevanz-Konzepts ausgelotet, hier sollen einige beispielhafte Hinweise auf praktische Implikationen genügen.

Queries verkürzen i.d.R. komplexere Interessen; das ist auch – aber nicht ausschließlich – der Neigung vieler Nutzer geschuldet, sich „Google-like“ auf einfache Terme in einem einzigen Suchfeld zu beschränken. Doch selbst, was einfach und gleich aussieht, muss nicht dasselbe bedeuten. Ein Beispiel: Eine Suchanfrage „Biodiversität“ kann zielen auf:

- das wichtigste Buch (Impact) zum Thema,
- das neueste Buch (Aktualität),
- das beste Buch (inhaltliche Relevanz für eine konkrete Forschungsfrage),
- kein Buch, sondern Zeitschriftenartikel (Dokumenttyp),
- möglichst alles zum Thema (Recall),
- konkrete Bücher: „Biodiversität: Wissenschaftliche Grundlagen und gesetzliche Relevanz“; auf eine Übersetzung der Publikation „Biodiversité“ von Ch. Lévêque und J.-C. Mounolu, oder der von W. Dudley „Biodiversity“, oder auch die Originalausgaben in englischer bzw. französischer Sprache.

Alle diese Fälle sind plausible Forschungsfragen, die Ausgangsquery ist immer „richtig“ und könnte nur im letzten Fall ohne weiteres präziser formuliert sein. Das Beispiel zeigt also, dass sich die differenzierten Interessen nicht so einfach in genaue Anfragen umsetzen lassen, jedenfalls nicht mit den Mitteln einer einfachen Suche, ohne entsprechende „Qualifier“, ohne Filtersetzung u. ä.

Wie lassen sich nun präzise Antworten auf diese – unterschiedlichen und in dieser Unterschiedlichkeit unformulierten – Fragen finden? Der oben genannten Reihe nach: Zitationen lassen sich zur Bestimmung des *Impacts* heranziehen, hochzitierte Publikationen können so höher gerankt werden. Das *Aktualitätsbedürfnis* lässt sich durch eine entsprechende Sortier- oder Neuigkeitenoption bedienen. Die spezifische inhaltliche *Relevanz*, die sehr individuell definiert wird, am besten durch möglichst weitgehende Informationen wie Inhaltsverzeichnis, Referenzen oder Volltext. Die Differenzierung nach *Publikationstyp* kann über Filtersetzung (in erweiterten Suchen) oder Eingrenzung über eine Facette realisiert werden. Der *Recall* kann durch unscharfe Suchen und automatische Termerweiterung (Stemming, Synonyme, Übersetzungen) erhöht werden. Automatische Übersetzung bzw. Adressierung von Mehrsprachigkeit wäre auch beim Auffinden der im letzten Punkt genannten konkreten Bücher hilfreich. All dies sind Formen von Suchunterstützung, die aber an ganz unterschiedlichen Stellen im Suchprozess ansetzen und zumindest zum Teil auf die Interaktion mit dem Suchenden angewiesen sind.

Gemeinsam ist den oben angeführten Fällen, dass es sich im Grunde um „Standardsituationen“ handelt; Aktualität, Wirkung, möglichst umfassende Informationen oder die Differenzierung nach Publikationstypen sind Suchinteressen, die der eingeübten Logik und Struktur akademischen Arbeitens entsprechen. Deshalb finden sich diese Parameter auch immer häufiger als Sortieroptionen und Verfeinerungsmöglichkeiten in wissenschaftlichen Informationsangeboten wieder.

„Spezialinteressen“ und individuelle Relevanzkriterien sind komplizierter zu bedienen, aber, wie im Abschnitt über „explorative Suchen“ angesprochen, genau das, was anspruchsvolle Nutzer im Wissenschaftsbereich als Anforderung an Informationsangebot und Retrieval herantragen. Auf Häufigkeiten beruhende Vorschläge helfen hier kaum weiter; vielversprechender ist das Bereitstellen vielfältiger und weitgehender Informationen, die eine individuelle Relevanzbewertung ermöglichen. Hierzu gehören neben den oben genannten, die sich auf das einzelne Dokument beziehen (Abstract, Referenzen, Volltext etc.), auch die Kontexte, die beispielsweise über Verlinkungen oder Anschlussuchen erschlossen werden können (z.B. Affiliationen, Koauthorschaften, Publikationsorte). Eine Auflistung von Kategorien, die zur Relevanzzuweisung beitragen, liefern Rieh und Belkin [40].

### *3.2. Suchpräzisierung („Vorher“-Selektion): Filter und Vokabular*

Eine gute Voraussetzung, um gute Informationen zu finden, ist die Formulierung einer guten Suchanfrage. Hierbei helfen auf das Suchinteresse abgestimmte Terme, die Nutzung einer qualitativ hochwertigen und differenzierten Erschließung, wie sie für Fachcontent häufig zur Verfügung steht, und die Möglichkeiten von erweiterten Suchen oder Experten-Suchen, die gezielt einzelne Felder adressieren und komplexe Queries erlauben. Eine solche Suchtaktik, bei der zunächst Thesaurus-Begriffe recherchiert, Labels kontrolliert, mehrere Felder gefüllt und mit Operatoren verknüpft werden sollen, ist aufwendig und dürfte im Informationshabitus der wenigsten Nutzer fest verankert sein. Trotzdem scheint sie lohnend vor allem in Fällen, in denen sie mit einem nachhaltigen Informationsinteresse verbunden ist, wenn beispielsweise personalisierte Alerts mit komplexen und genauen Suchanfragen hinterlegt werden.

Ein weiteres Mittel zur Präzisierung von Suchanfragen sind Filter, die von vornherein nur bestimmte Segmente des Suchraums zulassen. So bietet es sich an, wiederkehrende Sucheinstellungen in erweiterten Suchmasken zu speichern – wer beispielsweise regelmäßig ausschließlich Zeitschriftenpublikationen (oder Bilder) sucht, sich auf englischsprachiges Material beschränken möchte oder gezielt bestimmte Kollektionen durchsucht, kann den Ballast durch entsprechende Filtersetzungen minimieren. Wenn die rasche Verfügbarkeit ein Kriterium ist, ist es sinnvoll, die Suche auf digitale Volltexte zu begrenzen etc. Für das Design der Filtermöglichkeiten ist jeweils die fachliche Wissensstruktur entscheidend.

Wissenschaftliche Dokumente bedienen sich in der Regel einer Fachterminologie oder Nomenklatur, durch die sie im günstigsten Falle bereits auf textueller Ebene als eigener Suchraum definiert sind. In naturwissenschaftlichen und technischen Fächern ist das häufig so, für geistes- und sozialwissenschaftliche Fachbereiche gilt das leider nicht im gleichen Maße, wodurch das „vocabulary problem“ [41] hier in verschärftem Maße auftritt, nämlich mit einem Ballast an falschen Positiven. Auch in diesen Disziplinen wird selbstverständlich mit Fachbegriffen gearbeitet, diese sind aber keine künstlichen Prägungen, sondern beruhen auf Beschreibungen des Gegenstandsbereichs und haben oft Entsprechungen in der Alltagssprache. Das lässt sich leicht zeigen: Die

unterschiedlichen Bedeutungsumfänge des alltagssprachlichen und des fachsprachlichen Begriffs werden unmittelbar deutlich, wenn man mit demselben Term – z.B. „Ritual“ – einmal mit Google und einmal mit Google Scholar sucht. Die Informationen unterscheiden sich nicht nur qualitativ im Anspruchsniveau, sondern auch inhaltlich hinsichtlich dessen, was mit dem Begriff gemeint ist.

Ziel von Suchvorschlägen ist es generell, Kongruenzen zwischen Dokument bzw. Erschließung auf der einen und Query auf der anderen Seite zu erzeugen, um so genauere Treffer zu erzielen. Strohmaier et al. [42] nennen „query expansion“, „query substitution“, „query recommendation“ und „query refinement“ als Konzepte, die suchunterstützend zur Verbesserung von Anfragen beitragen. In Fachdatenbanken und Katalogen dienen kontrollierte Vokabulare und Klassifikationen zur Verfeinerung (z.B. über Facetten), sie können aber auch zur Unterstützung bei der Anfrageformulierung eingesetzt werden, um Erschließung und Query zu abzugleichen. Furnas et al. haben bereits vor einem Vierteljahrhundert festgestellt: „In information retrieval systems, the keywords that are assigned by indexers are often at odds with those tried by searchers.“ [41] Eine Möglichkeit, dieses Problem per Suchunterstützung zu adressieren, sind Termvorschläge, die aus kontrollierten Vokabularen wie Thesauri oder auch aus Indices und Normdaten kommen. Je spezifischer, differenzierter und näher am aktuellen Sprachgebrauch der Wissenschaftscommunity solche Erschließungsmittel sind, mit desto größerer Wahrscheinlichkeit werden sie akzeptiert und genutzt werden. Einen besonderen Mehrwert stellen dabei Vorschläge dar, die nicht auf rein lexikalischen Prinzipien beruhen, sondern solche, die semantische Nähe berücksichtigen (s. hierzu Abschnitt 5). Was die technische Seite anbelangt, sollten solche Empfehlungen *on the fly* angeboten werden, z. B. als Autocomplete-Service, der keine weitere Iteration erfordert.

### 3.3. Trefferpräsentation („Nachher“-Selektion): Facetten, Ranking

Gerade die Bevorzugung einfacher und schnell durchzuführender Suchanfragen, die große und unscharfe Treffermengen erzeugen, macht eine gut strukturierte Präsentation dieser Ergebnisse notwendig. Präzisierende Selektionen, die nicht im Vorhinein über das Suchinterface vorgenommen werden, können im Nachhinein über Facetten oder Sortieroptionen genutzt werden: Die Einschränkung nach Zeiträumen, Informationstypen, Kollektionen, Autoren, inhaltlichen Bereichen oder Verfügbarkeit können die Menge sinnvoll begrenzen. Rankings nach Aktualität, Impact oder Relevanz (s. 6.1) liefern Antworten auf unterschiedliche Suchinteressen.

Im Prinzip können alle Parameter, die zur „Vorher“-Selektion eingesetzt werden, auch herangezogen werden, um Ergebnislisten im Nachhinein einzuschränken. Das erzeugt zwar in gewisser Weise Doppelstrukturen, diese werden aber kaum als redundant empfunden werden, da sie unterschiedliche Suchstrategien unterstützen. Das Problem besteht eher darin, die Komplexität des Frontends auch in der Trefferpräsentation nicht ausufern zu lassen und das Angebot an Facetten und Re-Rankings auf häufig Nachgefragtes zu konzentrieren. Aggregierte und hierarchisierte Ansichten anzubieten, scheint ein guter Weg zu sein, um vielfältige und tiefe Informationen in übersichtliche Form zu bringen (s. Punkt 6.6).

Die folgenden Abschnitte beschreiben konkrete suchunterstützende Services auf den Ebenen Anfrageformulierung, Trefferlisten und Einzeldokument.

#### 4. Suchunterstützung in akademischen Suchmaschinen

Dieses Kapitel wird im weiteren Verlauf eine Auswahl von suchunterstützenden Instrumenten beschreiben und knüpft an die Beschreibung von wissenschaftlichen Suchmaschinen in Weinhold et al. [1] an. Die Erläuterungen der Beispiele orientieren sich an den Anforderungen, die heutzutage an Benutzerinterfaces bei Suchmaschinen gestellt werden und u. a. bei Hearst [43] beschrieben sind. Um die suchunterstützenden Maßnahmen zu illustrieren, wurden konkrete Beispiele aus aktuellen akademischen Suchmaschinen ausgewählt. Die Auswahl wurde nach einer Auswertung von vielen in Frage kommenden akademischen Suchmaschinen und deren suchunterstützenden Features getroffen und soll ein möglichst breites Spektrum wiedergeben. Die in diesem Artikel erwähnten Suchmaschinen und die darin angebotenen Informationsarten und fachlichen Domänen werden in Tabelle 1 aufgezählt.

**Tabelle 1.** Exemplarische Auswahl der untersuchten akademischen Suchmaschinen

<b>Name</b>	<b>Domäne</b>	<b>Umfang (03/2013)</b>	<b>Besonderheit</b>
<b>BASE</b> base-search.net	verschiedene Disziplinen	ca. 42 Mio. Texte, Aufsätze, Bücher, Reports, Bilder etc.	Aggregator von OA-Repositoryn, Volltextsuche
<b>Web of Knowledge</b> webofknowledge.com	verschiedene Disziplinen	ca. 12.000 ausgewertete Zeitschriften, ca. 150.000 Proceedings	Zitationen, Abstracts
<b>PubMed</b> Pubmed.gov	Medizin, Biomedizin	ca. 22 Mio. Informationseinheiten	Zitationen, Abstracts
<b>Scopus</b> scopus.com	verschiedene Disziplinen	ca. 49 Mio. Zeitschriftenartikel	Zitationen, Abstracts
<b>Science Direct</b> sciencedirect.com	verschiedene Disziplinen	ca. 12 Mio. verlagseigene Artikel	Verlagssystem
<b>Springer Link</b> link.springer.com	verschiedene Disziplinen	ca. 7 Mio. verlagseigene Artikel, Buchkapitel	Verlagssystem
<b>Sowiport</b> sowiport.de	Sozialwissenschaften	ca. 7 Mio. Literaturnachweise, Projekte, Studien etc.	Zitationen, Abstracts, Volltextsuche, redaktionelle Inhalte
<b>Google Scholar</b> scholar.google.de	verschiedene Disziplinen	Artikel, Patente	Volltextsuche, Profile
<b>Deutsche Digitale Bibliothek</b> deutsche-digitale-bibliothek.de	Geisteswissenschaften, Kulturgeschichte	ca. 6 Mio. Bilder, Texte, Ton-, Videoaufnahmen	nicht-textuelle Informationstypen
<b>Europeana</b> europeana.eu	Geisteswissenschaften, Kulturgeschichte	ca. 25 Mio. Bilder, Texte, Ton-, Video- und 3D-Aufnahmen	nicht-textuelle Informationstypen

Die Beschreibung der suchunterstützenden Instrumente erfolgt detailliert in den nachfolgenden Abschnitten. Wir ordnen diese jeweils einer Phase im Suchprozess zu. Der Prozess der Suche beginnt mit der Formulierung oder Auswahl einer Suchanfrage und geht über die Selektion eines Dokuments aus einer Trefferliste in eine detaillierte Dokumentenansicht über. Zunächst werden in Abschnitt 5 Instrumente erläutert, die zu Beginn eines Suchvorhabens dem Benutzer auf der Startseite angeboten werden. In Abschnitt 6 werden suchunterstützende Instrumente gezeigt, die vom Benutzer verwendet werden können, wenn ihm erste Ergebnisse seiner Recherche vorliegen. Dies erfolgt in allen genannten Suchmaschinen in Form einer Trefferliste. Abschnitt 7 zeigt Instrumente, die dem Benutzer angeboten werden, wenn er ein bestimmtes Dokument aus der Trefferliste ausgewählt hat und eine weitere Suche daran anknüpfen möchte.

## **5. Suchunterstützung bei der Anfrageformulierung**

Eines der klassischen Probleme im Information Retrieval ist die Vagheit, die zwischen einer Anfrage eines Benutzers und den Begriffen, die ein Dokument beschreiben (in unserem Kontext i. d. R. Deskriptoren eines Thesaurus), besteht. Konkret bedeutet dies, dass ein Benutzer für eine beliebige Fragestellung häufig andere Suchterme verwendet, als derjenige, der die Inhalte eines Dokuments verfasst (Autor) oder derjenige, der die Inhaltserschließung für entsprechende Dokumente vornimmt (Indexierer) [vgl. 41]. Für die Suche in großen Suchräumen kann das z. T. sehr negative Auswirkungen wie z. B. wenige oder keine Treffer oder zu viele und irreführende Treffer (z. B. Homonymproblem) haben. Dieses Problem verschärft sich in akademischen Suchmaschinen, die Dokumente aus unterschiedlich erschlossenen (heterogenen) Datenbanken enthalten (wie z. B. sowiport mit insgesamt 19 Datenbanken und 10 unterschiedlichen Erschließungssystemen).

Gegenüber einfachen Anfrageexpandierungs- oder Termvervollständigungsverfahren (z. B. Googles Vorschlags-Feature) haben Vorschlagssysteme in akademischen Suchmaschinen den entscheidenden Vorteil, dass Terme aus dem bei der Inhaltserschließung verwendeten kontrollierten Vokabular vorgeschlagen werden können. Dies erhöht die Erfolgchancen, relevante Dokumente zu einem bestimmten Konzept zu finden.

Die Vagheit zwischen Anfrage- und Dokumentebene kann durch unterschiedliche Verfahren der Termerweiterung behandelt werden. Dies konnte in folgenden Studien nachgewiesen werden. Mayr und Petras [44] haben in einem kontrollierten IR-Experiment belegt, dass Verfahren, die unterschiedliche Erschließungsvokabulare auf Basis von semantischer Äquivalenz verknüpfen (sog. intellektuell erstellte Crosskonkordanzen), sehr positive Ergebnisse bezogen auf die Trefferrelevanz und -vollständigkeit haben. Hienert et al. [45] haben in einer umfangreichen Benutzerstudie (n=4000 Benutzer) mit der akademischen Suchmaschine sowiport gezeigt, dass vier ähnliche Termempfehlungsdienste sehr unterschiedlich von den Benutzern aufgenommen werden. Die Studie belegt, dass derjenige Dienst die beste Akzeptanz aufweist, der alternative Suchterme bei der Anfrageformulierung vorschlägt, die über die einfache Autocomplete-Funktionalität hinausgehen. Lücke et al. [46] zeigen in einem kontrollierten IR-Experiment, dass disziplinspezifische kontrollierte Termvorschläge auf Grundlage von spezifisch trainierten Dokumentkorpora aufgrund

der höheren fachlichen Spezifität der Begriffe eine höhere Präzision bei der automatischen Suchtermexpansion erreichen können.

Die meisten Suchmaschinen bieten den Nutzern bereits auf der Startseite zahlreiche Funktionen an, die der Unterstützung der Suche dienen. Fast immer sind dies ein oder mehrere Eingabefelder für die Suchanfrage in Form eines Suchformulars oder, falls der Benutzer bereits Anfragen gestellt hat, in Form einer Suchhistorie, in der die bisherigen Anfragen aufgelistet werden.

### *5.1. Suchformulare*

Sucheingabefelder oder Suchformulare haben sich für die Formulierung einer Anfrage gegenüber vordefinierten Browsing-Zugängen in allen allgemein bekannten und den am meisten benutzten Suchmaschinen durchgesetzt. Deshalb ist es sehr wichtig, dass diese verständlich und einfach zu benutzen sind, schließlich muss es dem Benutzer möglich sein, mit den Funktionalitäten der Suchformulare sein Informationsbedürfnis auszudrücken. Dem Benutzer können unterschiedliche Arten von Suchformularen angeboten werden, die sich im Umfang der abgefragten Informationen unterscheiden. Die einfachste Form zeigt dem Nutzer ein Suchfeld, über das er zu seinen gewünschten Informationen gelangt. In diesem Feld können ein oder mehrere Suchbegriffe eingegeben werden, mit denen im gesamten Datenbestand und allen bzw. den für die konkrete Suche wichtigen Feldern gesucht wird. Bei mehreren Suchbegriffen wird in der Regel eine AND- oder eine OR-Verknüpfung angewendet.

Über einfache Suchanfragen hinaus gehen die Möglichkeiten, die in Formularen von „erweiterten Suchen“ angegeben werden können. Die erweiterten Suchen bieten dem Nutzer die Möglichkeit, die Anfragen genauer zu spezifizieren, d. h. er kann z. B. festlegen, in welchen Attributen nach den gewünschten Begriffen gesucht werden soll oder welche Datenbestände, Zeiträume oder Informationsarten berücksichtigt werden sollen.

So bietet die Suchmaschine BASE dem Benutzer in der erweiterten Suche beispielsweise eine bestimmte Anzahl von Suchfeldern, denen ein bestimmtes Informationsfeld zugeordnet ist, welches mit diesem Suchformularfeld durchsucht werden soll. Dabei ist die Zuordnung, welches Attribut in einem Feld durchsuchbar werden soll, über eine Auswahlliste änderbar. Einige Suchmaschinen, wie z. B. die Deutsche Digitale Bibliothek (s. Abb. 1), bieten dem Nutzer eine dynamisch erweiterbare Anzahl von Feldern für komplexe Suchanfragen an. Zudem kann die Art der Verknüpfung (ODER, UND) zwischen den Feldern geändert werden. Die Deutsche Digitale Bibliothek bietet zudem die Möglichkeit, innerhalb der erweiterten Suche Suchgruppen zu definieren. Einige wenige Suchmaschinen wie z.B. PubMed mit dem „Pubmed Advanced Search Builder“ bieten eine Expertensuche, in der man frei über ein mehrzeiliges Textfeld hochkomplexe Anfragen aus booleschen Operatoren und Feldbezeichnungen definieren kann (s. Abb. 2). Diese richtet sich an erfahrene Nutzer, die in der Lage sind, komplexe Anfragen über die Syntax zu definieren, um darüber an sehr genaue Ergebnisse zu gelangen.



Verknüpfung der Suchgruppen ODER ▾

Verknüpfung der Suchfelder ODER ▾ ✕

Alles durchsuchen ▾  Einer der Begriffe ▾ ✕

Alles durchsuchen ▾  Einer der Begriffe ▾ ✕ +

Verknüpfung der Suchfelder ODER ▾ ✕

Alles durchsuchen ▾  Einer der Begriffe ▾ ✕

Alles durchsuchen ▾  Einer der Begriffe ▾ ✕ +

Suchgruppe hinzufügen +

Abb. 1: Erweiterte Suche der Deutschen Digitalen Bibliothek (Definition von Suchgruppen)

((EHR OR "patient summary" OR "patient record")) OR ((telemedicine OR telehealth OR e-health) AND patient)) AND ("2006"[Date - Publication] : "2013"[Date - Publication])) AND evaluation[Text Word]

Abb. 2: Suchsyntax im PubMed Advanced Search Builder

## 5.2. Suchhistorie

Ein weiteres wichtiges Feature ist die Anzeige einer Suchhistorie, die bei einigen Suchmaschinen neben dem bekannten Suchformular bereits auf der Startseite angezeigt wird. Benutzer von Suchmaschinen stellen oft gleiche Suchanfragen, ein erneutes Aufrufen von bereits durchgeführten Aktionen kann die Produktivität bei der Suche erhöhen [43]. Ein Beispiel für eine Suchhistorie lässt sich bei Scopus finden (s. Abb. 3). Die Suchhistorie führt dabei die 50 zuletzt ausgeführten Anfragen der aktuellen Suchsession auf. Der Nutzer kann diese für weitere Suchanfragen aufgreifen, die sich auch über die Funktion „Combine queries“ mittels boolescher Operatoren verknüpfen lassen. Hat der Benutzer eine Personalisierung vorgenommen, lassen sich Suchanfragen auch dauerhaft speichern. Bei Scopus besteht außerdem die Möglichkeit, bereits ausgeführte Suchen zu editieren bzw. sich per RSS oder Alert über Änderungen informieren zu lassen. Bei der RSS-Funktion wird dem Nutzer speziell zu seiner Suche ein RSS-konformes XML angeboten, welches mittels RSS-Readern in periodischen Abständen abgerufen werden kann und so über Neuzugänge bzw. Änderungen informiert. Ähnlich verhält es sich mit einer Alert-Funktion, hier erfolgt die Benachrichtigung über Änderungen in gespeicherten Suchergebnissen meist per Email.

Search history <span style="float: right;">Hide ▾</span>						
Combine queries <input type="text" value="e.g. (#1 AND #2) AND NU1 #3"/> <input type="button" value="Search"/> <a href="#">? Combining queries</a>						
Search	Results	Set feed	Set alert	Save	Edit	Delete
1 TITLE-ABS-KEY(diabetes)	530,689					

Abb. 3: Suchhistorie von Scopus

## 6. Suchunterstützung auf Trefferlistenebene

Sobald eine Suchanfrage gestellt wurde, erfolgt die Darstellung der Ergebnisse in Form von textuellen Trefferlisten, die komprimiert eine kurze Zusammenfassung mit den wichtigsten Informationen über die gefundenen Dokumente geben. Bei Literatureinträgen umfasst die Schnittmenge zwischen allen vorgestellten Suchmaschinen Informationen über den Titel, die Autoren und dem Erscheinungsjahr. Oft wird noch ein Abschnitt aus dem Abstract angezeigt, sofern der Inhalt Angaben zu gesuchten Begriffen enthält. Es gibt noch weitere Darstellungsformen, auf diese soll aber in diesem Abschnitt nicht eingegangen werden.

### 6.1. Ranking

Trotz großer Dokumentmengen für datenbankübergreifende Literaturrecherchen erwarten akademische Nutzer heute einen möglichst hohen Anteil an relevanten und qualitativen Dokumenten in den Trefferergebnissen. Insbesondere die Reihenfolge und Struktur der gelisteten Ergebnisse (sog. Ranking) spielt, neben dem direkten Volltextzugriff auf die Dokumente, inzwischen eine entscheidende Rolle beim Design von Suchsystemen.

Die neueren technologischen Entwicklungen des Web Information Retrieval (IR), wie sie Google implementiert, werten insbesondere frei zugängliche Dokumente mit ihrer gesamten Text- und Linkinformation automatisch aus. Diese Verfahren sind vor allem deshalb erfolgreich, weil sie Ergebnislisten nach Relevanz gerankt darstellen und einfach und schnell zu recherchieren sind. Die qualitativen Verfahren der traditionellen Informationsanbieter (z. B. WoK oder sowiport) hingegen zeigen genau bei diesen Punkten (Ranking, Einfachheit und Volltextzugriff) Schwächen, überzeugen aber vor allem durch ihre Stringenz, in diesem Fall die selektive Aufnahme von qualitätsgeprüften Dokumenten in das System und die inhaltliche Erschließung der Dokumente. Aufgrund der Beschaffenheit der Dokumente in klassischen Informationssystemen greifen die Ranking-Verfahren des Web-IR genauso wenig wie die herkömmlichen textstatistischen Information Retrieval-Verfahren. Neue Ansätze wie z. B. der Einsatz von Autorenzentralitätsmaßen beim Ranking, Zeitschriftenproduktivität und Autorenproduktivität (s. sowiport in Abb. 4) bei Mutschke et al. [47] zeigen, dass die spezifische Struktur der Literaturnachweise und die Qualität der Daten für alternative Mehrwertdienste gewinnbringend eingesetzt werden kann (s. auch [57]).

Abgegrenzt wird Ranking oder Relevance Ranking von sog. Sortierungen z. B. nach dem Erscheinungsjahr der Publikation, obwohl hier die Grenze zu „nach inhaltlicher Relevanz“ gerankten Listen konzeptuell nicht sauber zu ziehen ist. Das Ranking von Dokumenten führt letztlich dazu, dass sich die Benutzer solcher Suchsysteme fokussiert mit den oberen Treffermengen (top-gerankt) eines Suchergebnisses beschäftigen. Der mittlere und untere Bereich eines Suchergebnisses wird häufig nicht mehr in Betracht bezogen. Bei großen Treffermengen ist eine vollständige Rezeption eines Suchergebnisses ohnehin nicht mehr möglich. Aufgrund der Vielzahl an relevanten und verfügbaren Informationsquellen ist es daher notwendig, Kernbereiche in den Suchräumen zu identifizieren und diese anschließend

dem Nutzer hervorgehoben zu präsentieren. Die Kernbereichsbildung kann zum einen über statistische Ranking-Verfahren erfolgen, zum anderen aber auf Grundlage der den Suchräumen inhärenten Strukturen (siehe dazu Mayr [48]).

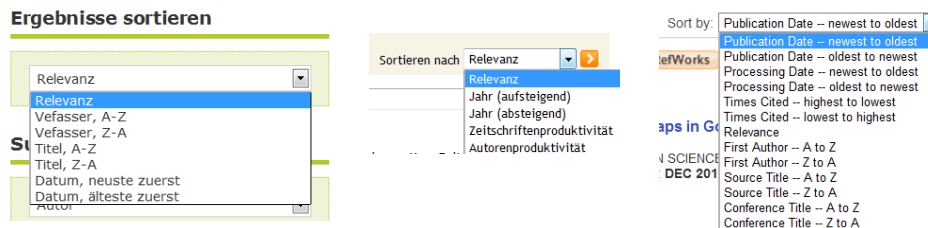


Abb. 4: Auswahllisten des Ranking aus BASE, sowiport und Web of Knowledge

## 6.2. Navigation / Facetten

Mit Hilfe eines Navigationsbereiches bzw. von Facetten ist der Benutzer in der Lage, sich einen zusammenfassenden Überblick über das Gesamtergebnis seiner Suchanfrage geben zu lassen, während die Trefferliste selbst zunächst nur die ersten 10 oder 20 Treffer umfasst und weitere Treffer erst über einen Abruf von zusätzlichen Treffern bietet. Die im Navigationsbereich dargestellten Informationen sind kategorisiert und zeigen Werte an, die in den jeweiligen Kategorien in der gesamten Ergebnismenge enthalten sind. Diese Einträge lassen sich für eine weitere Eingrenzung des Ergebnisses nutzen. Über die Navigation bekommt der Nutzer Schlüsselwörter aufgelistet, die ihm ohne diese Funktion nicht präsent sind und er kann diese bequem auswählen [49] [50]. Innerhalb der Gruppierung werden die Einträge anhand ihrer Häufigkeit geordnet. Facetten lassen sich in allen untersuchten Suchmaschinen finden.

Besonders hervorzuheben ist die Funktion beim Web of Knowledge (s. Abb. 5). Dort werden dem Nutzer 13 Kategorien (z. B. Web of Science Kategorie, Dokumenttyp, Autor) zu seinem Ergebnis angezeigt. Die Einträge in den Kategorien werden nur durch eine individuelle Auswahl sichtbar und umfassen zunächst nur die fünf häufigsten Einträge. Jedem Eintrag ist eine Checkbox vorangestellt, die beim Auswählen dieses Eintrages das Ergebnis auf die gewählte Ausprägung eingrenzt. Beim Web of Knowledge lassen sich über die Checkboxes mehrere Einträge aus dem Navigationsbereich gleichzeitig auswählen. Ein weiteres Feature ist die Möglichkeit, über ein Suchfeld innerhalb der Navigationseinträge nach bestimmten Werten zu suchen (s. Abb. 5).



Abb. 5: Ausschnitt aus dem Facettenbereich des Web of Knowledge

### 6.3. Term Highlighting

Das Term Highlighting hebt die gesuchten Terme innerhalb der Ergebnisdarstellung visuell hervor. Dies geschieht z. B. typographisch durch abweichende Schriftstärke oder -größe oder durch einen geänderten farblichen Hintergrund. Die Trefferdarstellung des Web of Knowledge setzt die Hervorhebung durch einen andersfarbigen Hintergrund um (s. Abb. 5). Durch das Hervorheben wird der Blick des Nutzers direkt auf den Suchbegriff geleitet und ihm wird zusätzlich gezeigt, warum das jeweilige Suchergebnis in der Trefferliste enthalten ist und ob dieses für ihn relevant ist [51].

### 6.4. Related Searches

Unter „Related Searches“ wird bei Pubmed eine Funktion verstanden, die dem Benutzer weitere, zur bisherigen Suche inhaltlich relevante Suchterme und -anfragen anbietet (s. Abb. 6). Der Benutzer bekommt so weitere und für ihn interessante Suchanfragen angeboten, und kann diese in Form einer Vorschlagsliste für eine weitere Suche auswählen. Bei Pubmed werden dem Nutzer die häufigsten Suchanfragen anderer Nutzer auf der rechten Seite neben der Trefferliste angezeigt, die den oder die gesuchten Begriffe enthalten. Der gewählte Suchbegriff wird, wie es auch bei der Hervorhebung der Suchbegriffe in der Trefferanzeige geschieht, durch einen fetten Schriftschnitt herausgestellt.

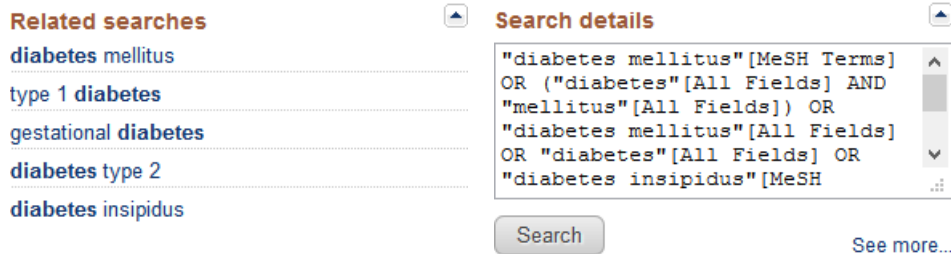


Abb. 6: Related Searches und Search Details bei Pubmed

### 6.5. Search Details

Pubmed enthält im rechten Bereich neben der Trefferlistendarstellung eine Funktion, welche dem Benutzer die genaue Syntax der Suchanfrage präsentiert (s. Abb. 6). Sehr genau wird dort angezeigt, in welcher Entität nach welchem Begriff gesucht worden ist. Die Anzeige kann direkt dazu genutzt werden, die Suchanfrage zu modifizieren und erneut zu stellen.

### 6.6. Aggregation unterschiedlicher Informationen

Zunehmend gibt es Suchmaschinen, die eine Aggregation von verschiedenen Informationen in der Trefferpräsentation anbieten, so wie sie auch in [52] als „Universal Search“ oder in [53] als „Aggregated Search“ beschrieben werden. Dabei werden nach einer Suche in einem heterogenen Bestand der Suchmaschine

unterschiedliche Informationstypen oder Quellen dem Nutzer gemeinsam in einer Ansicht präsentiert, die bisher aus gesonderten Ansichten stammen. Verbreitet ist dieses Verfahren bei den großen bekannten Websuchmaschinen wie z.B. Google oder Yahoo!, bei denen die unterschiedlichen Informationstypen als Webinformationen, Bilder, Videos, News, Landkarten usw. definiert werden und Informationen aus diesen Typen strukturiert gemeinsam im Resultat einer Suche angezeigt werden. Der Nutzer besitzt oft nicht die Kenntnis, welche Informationsarten für ihn interessant sein können, daher bestimmt das System anhand einiger Verfahren, welche Auswahl an Informationsquellen für die Suche einbezogen werden sollte, welche Ergebnisse daraus entnommen werden und wie diese Ergebnisse für den Benutzer anzuordnen sind. Ein Merkmal für die Bestimmung der Auswahl kann ein Keyword wie z.B. „Stadtplan“ in der Suchanfrage sein, an dem die Suchmaschine erkennen kann, dass bei einer Suche nach „Kölner Stadtplan“ im Ergebnis Anzeigen des Stadtplans als Bilder oder Landkarten enthalten sein sollten. Eine umfangreiche Erläuterung über diese Aspekte finden sich in Arguello et al. [54].

Dieses Prinzip der Aggregation lässt sich auch auf wissenschaftliche Suchmaschinen übertragen, wenn unterschiedliche Informationsarten von deren Bestand abgedeckt werden. Einen Ansatz verfolgt Pubmed mit PMC Images Search (s. Abb. 7). Zu den eigentlichen Textzusammenfassungen der Trefferliste werden in einem kleinen Kasten auf der rechten Seite Bilder angeboten, die ebenfalls der Suche entsprechen. Die Anzeige des Bildbereichs erfolgt nur, wenn mindestens vier Bilder verfügbar und weitere Aspekte erfüllt sind.

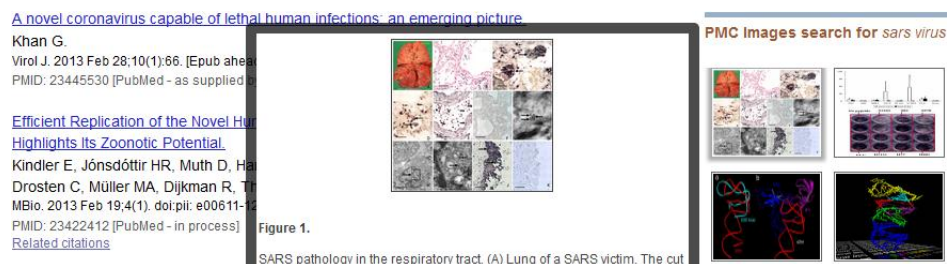


Abb. 7: PMC Images Search in der Seitenleiste der Trefferliste von Pubmed

## 7. Suchunterstützung auf Dokumentebene

Nach der Suche und der Begutachtung der erhaltenen Resultate in Form von kurzen Zusammenfassungen auf Trefferseiten kann sich der Nutzer durch Anklicken eines Treffers für eine Detailansicht des Dokuments entscheiden. Diese Detailansicht erscheint entweder eingebettet in die Listendarstellung, wie es z. B. bei sowiport angeboten wird, oder isoliert von weiteren Ergebnissen auf einer eigenen Seite. In der Detailansicht werden dem Benutzer alle zu dem gewählten Dokument gespeicherten Informationen angezeigt. Der Benutzer kann sich ausführlich über das Dokument informieren und innerhalb der Ansicht Instrumente für weiterführende Dokumente nutzen oder die Informationen für eine weitere Verwendung speichern oder exportieren.

Neben den in einer Dokumentansicht gezeigten Informationen stehen bei den jeweiligen Suchmaschinen oft weitere Informationen zur Verfügung, die für den Nutzer

im Kontext des gewählten Dokuments interessant sein können. Zu den einzelnen Informationen eines Dokuments können Links hinterlegt werden, die zu weiteren Dokumenten führen, die ebenfalls hilfreich sein könnten. Das Angebot von Verlinkungen führt so zu einem explorativen Suchprozess, bei dem der Nutzer zu immer weiteren Dokumenten gelangen kann.

Die Verlinkungen zu weiteren Dokumenten können auf unterschiedliche Weise implementiert werden. Bei den meisten Suchmaschinen wird dem Nutzer die Möglichkeit einer Anschlussuche gegeben. Dabei werden Informationen aus bestimmten Feldern für weitere Suchen verwendet. Im Gegensatz zur Auswahl einer Facette aus dem Navigationsbereich, wird hier nicht die Ergebnismenge der vorher durchgeführten Suche weiter reduziert, sondern es wird eine neue Suche ausgeführt, die nur die gewählte Ausprägung in dem zugehörigen Feld besitzt. Oft wird diese Technik bei Informationen über Autoren oder Feldern mit kontrolliertem Vokabular verwendet.

Eine weitere Anwendung von Verlinkungen zeigt auf bestimmte Dokumente, die mit dem angezeigten Dokument in einer inhaltlichen Verbindung stehen. Die Art der Verbindung kann hierbei unterschiedlich begründet sein. Dokumente entstammen z. B. einem übergeordneten Werk oder einem gemeinsamen Projekt, da ist es sehr hilfreich, wenn diese auch untereinander verlinkt sind. Dies gilt ebenfalls für Referenzen oder Zitationen von Dokumenten, wenn über einen Link die Rezeption eines Dokuments (der Zitationsverlauf) verfolgt werden kann [56].

In einem gesonderten Bereich gibt es in einigen akademischen Suchmaschinen einen abgegrenzten Bereich in der Dokumentansicht, worin dem Nutzer „verwandte“ Dokumente vorgeschlagen werden. Dieser Bereich ist von der Anzeige des Dokuments separiert, da die dort gezeigten Informationen nicht dem Inhalt des gezeigten Dokuments entstammen, sondern über Ähnlichkeitsmaße vom System ermittelt worden sind. Dies kann z.B. über probabilistische Verfahren gelöst werden, die die Topics miteinander vergleichen [55].

Sowiport bietet in seiner Detaildarstellung von Dokumenten die Möglichkeit, bei Autoren und Schlagwortangaben neue Suchen zu starten und im Referenz- und Zitationsabschnitt Links zu den aufgeführten Dokumenten zu folgen (s. Abb. 8).

Revisiting labeling theory and health: A commentary on Aronowitz [▲](#)  
 Autor: [Nathanson, Constance](#)  
 Erscheinungsjahr: 2008; Informationstyp: Literatur; Dokumenttyp: Zeitschriftenaufsatz  
 Datenbank: Sociological Abstracts (CSA)  
 Referenzen: 7; Zitiert von (Beta-Version): 2

<b>Schlagwörter</b>	<a href="#">Health</a> , <a href="#">Social Factors</a> , <a href="#">Social Constructionism</a> , <a href="#">Diseases</a> , <a href="#">Epidemiology</a> , <a href="#">Labeling</a>
<b>Klassifikation</b>	sociology of health and medicine; sociology of medicine & health care
<b>Sprache</b>	englisch
<b>Referenzen</b>	Aronowitz, R (2008): Framing disease: An underappreciated mechanism for the social patterning of health <i>Social Science &amp; Medicine</i> 67, 1, 1-9.
<b>Zitiert von</b>	Aronowitz, Robert (2008): Rejoinder to commentaries on 'Framing disease: an underappreciated mechanism for the social patterning of health'

Abb. 8: Detailansicht in sowiport mit Links zu weiteren Suchen sowie Referenzen und Zitationen

## **Zusammenfassung und Ausblick**

Das Ziel des Kapitels war es, Entwicklungstendenzen und praktische Services im Bereich der Suchunterstützung für wissenschaftliche Suchmaschinen aufzuzeigen. Die Besonderheiten wissenschaftlichen Recherchierens – Weiterverarbeitung der Ergebnisse im Forschungsprozess, Bedeutung explorativer Suchen, Wichtigkeit eigener Bewertungskriterien vor allgemeinen Empfehlungen, hoher Qualitätsanspruch – gehen über das Frage-Antwort-Paradigma hinaus, und sie lassen sich mit den Möglichkeiten allgemeiner Suchmaschinen oft nicht hinreichend bedienen.

Eine grundsätzliche Strategie zur Verbesserung von Retrievalergebnissen ist die interaktive Nutzung des Expertenwissens der User: mit Filtermöglichkeiten und Vorschlagsservices bei der Anfrageformulierung, mit Navigationsmöglichkeiten, Facettierung und aggregierten Ansichten bei der Trefferpräsentation, mit dem Angebot von weiterführenden Kontextinformationen auf Ebene des einzelnen Dokuments. Voraussetzung dafür, dass solche Erweiterungen angenommen werden, ist, dass sie der intellektuellen Struktur und der Arbeitspraxis des Felds, das sie bedienen, auch entsprechen bzw. diese sinnvoll weiterentwickeln. Um technische Laborentwicklungen in die Praxistauglichkeit entlassen zu können, ist es hilfreich, das Suchverhalten der konkreten Nutzergruppen zu erheben und in Services zu ‚übersetzen‘. Für das wissenschaftliche Rechercheverhalten gibt es eine ganze Reihe von Studien und Erfahrungen; trotzdem müssen sich Dienste immer wieder daraufhin überprüfen lassen, ob und wie sie Suchen einfacher und besser machen, ohne selbst erklärungsbedürftig zu sein.

Zu den Desideraten zählen suchunterstützende Tools, die die nahtlose Recherchierbarkeit und Verfügbarkeit von Informationen gewährleisten, dabei auch heterogene Informationstypen bewältigen und unterschiedliche Quellen abfragen. Die Ergebnisse von Meta- und integrierten Suchen können z.B. in aggregierten Darstellungsformen ausgegeben werden. Wichtig ist auch die komfortable Nutzung und Weiterverarbeitung von Rechercheergebnissen im akademischen Workflow – als Zitationsexport, Verfügbarkeitsabfrage in Bibliotheken, Download mit Hinweisen zu Nutzungsrechten etc. Explorative Suchstrategien werden effektiv ermöglicht durch Verlinkungen, die konkrete Forschungskontexte abbilden oder Bezüge herstellen: zum Beispiel zwischen Publikationen, erhobenen Daten, beteiligten Personen, Forschungsgruppen, Organisationen, wissenschaftlichen Konferenzen.

## Anhang

**Tabelle 2.** Feature-Übersicht akademischer Suchmaschinen, Teil 1

	<b>BASE</b>	<b>Web of Knowledge</b>	<b>PubMed</b>	<b>Scopus</b>	<b>Science Direct</b>
<b>Suchtermvorschlag / -erweiterung</b>	zusätzliche Wortformen, multilinguale Synonyme		Autocomplete, Termerweiterung		
<b>Suchformular</b>	einfach, erweitert	erweitert, fortgeschritten, Autoren, Referenzen	einfach, erweitert	erweitert, fortgeschritten, Autoren, Affiliationen	einfach, erweitert
<b>Suchhistorie mit Funktionen</b>	Chronik, RSS, Alerting	Chronik, RSS, Alerting	Chronik, Editierung	Chronik, RSS, Alerting, Editierung, Verknüpfung	Chronik, RSS
<b>Ranking</b>	Relevanz, Verfasser, Veröffentlichungsdatum	Aktualisierungsdatum, Veröffentlichungsdatum, Anzahl von Zitationen, Relevanz, Erstautor, Quellentitel, Konferenztitel	Aktualisierungsdatum, Veröffentlichungsdatum, Erstautor, letztgenannter Autor, Zeitschriftentitel	Relevanz, Veröffentlichungsdatum, Anzahl von Zitationen, Erstautor, Quellentitel	Relevanz, Veröffentlichungsdatum
<b>Navigation / Facetten</b>	Autor, Schlagwort, DDC, Jahr, Quelle, Sprache, Dokumentart, Zugang	Kategorie, Dokumentart, Forschungsbereich, Quelle, Person, Gruppe, Verfasser, Datum, Organisation, Geldgeber, Sprache, Land; mehrfache Selektion; Filtern nach eigenen Begriffen	Dokumentart, Volltext Verfügbarkeit, Datum, Spezies	Datum, Person, Schlagwort, Dokumentart, Quelle, Affiliation, Land, Sprache; mehrfache Selektion; Filtern nach eigenen Begriffen	Dokumentart, Quelle, Thema, Datum; mehrfache Selektion; Filtern nach eigenen Begriffen
<b>Term Highlighting</b>	fett geschrieben	farblich hinterlegt	fett geschrieben		fett geschrieben
<b>Anschliessende Suche</b>	Autor, DDC	Autor	Autor	Autor	Autor
<b>Verlinkung</b>	Datenlieferant, GoogleScholar	Related Documents		Zitationen, Related Documents	Related Documents, Related References



**Tabelle 2.** Feature-Übersicht akademischer Suchmaschinen, Teil 2

	<b>Springer Link</b>	<b>Sowiport</b>	<b>Google Scholar</b>	<b>Deutsche Digitale Bibliothek</b>	<b>Europeana</b>
<b>Suchtermvorschlag / -erweiterung</b>	Autocomplete	Autocomplete, Termerweiterung, STR			Autocomplete, Rechtschreibprüfung
<b>Suchformular</b>	einfach	einfach, erweitert	einfach, erweitert	einfach, erweitert (Suchgruppen, beliebig erweiterbar)	einfach
<b>Suchhistorie mit Funktionen</b>	RSS	Chronik, Abspeichern in Ordern	Alerting, Meine Updates (Vorschläge über Benutzerprofil)		Chronik
<b>Ranking</b>	Relevanz, Aktualisierungsdatum	Veröffentlichungsdatum, Relevanz, Zeitschriften-Produktivität, Autoren-Produktivität	Relevanz (berücksichtigt auch Anzahl von Zitationen), Veröffentlichungsdatum	Relevanz, Titel	Relevanz
<b>Navigation / Facetten</b>	Dokumentart, Disziplin, Unterdisziplin, Quelle, Sprache	Informationstyp, Anbieter, Person, Schlagwort, Datum	Datum	Datum, Ort, Person, Stichwort, Sprache, Medientyp, Sparte, Anbieter; Filtern nach eigenen Begriffen	Medientyp, Sprache, Datum, Land, Copyright, Anbieter
<b>Term Highlighting</b>	farblich hinterlegt		fettgeschrieben	fettgeschrieben	
<b>Anschliessende Suche</b>		Autor, Schlagwort			Autor, Subjekt, Jahr, Datum, Anbieter, Ähnlicher Inhalt
<b>Verlinkung</b>	Related documents	Referenzen, Zitationen, Related Documents, GoogleScholar	Related Documents, Zitationen		

## Literatur

- [1] Weinhold, Th.; Bekavac, B.; Schneider, G.; Bauer, L.; Böller, N. (2011): Wissenschaftliche Suchmaschinen – Übersicht, Technologien, Funktionen und Vergleich. In Lewandowski, D. (Hrsg.) (2011): *Handbuch Internet-Suchmaschinen 2. Neue Entwicklungen in der Web-Suche*. Heidelberg: Aka, 141–177.
- [2] Pieper, D.; Wolf, S. (2009): Wissenschaftliche Dokumente in Suchmaschinen. In: Lewandowski, D. (2009): *Handbuch Internet-Suchmaschinen: Nutzerorientierung in Wissenschaft und Praxis*. Heidelberg: Aka, 356-374.
- [3] Lewandowski, D. (2011): Query Understanding. In: ders. (Hrsg.) (2011): *Handbuch Internet-Suchmaschinen 2. Neue Entwicklungen in der Web-Suche*. Heidelberg: Aka, 55–75.
- [4] Quirnbach, S. (2011): Usability und User Experience in Suchmaschinen. In: Lewandowski, D. (Hrsg.) (2011): *Handbuch Internet-Suchmaschinen 2. Neue Entwicklungen in der Web-Suche*. Heidelberg: Aka, 229–263.
- [5] Abschnitt II in Lewandowski, D. (Hrsg.) (2011): *Handbuch Internet-Suchmaschinen 2. Neue Entwicklungen in der Web-Suche*. Heidelberg: Aka.
- [6] Ingwersen, P.; Järvelin, K (2005): *The turn: Integration of information seeking and retrieval in context*. Dordrecht: Springer.
- [7] Jansen, B. J.; Rieh, S.Y. (2010): The seventeen theoretical constructs of information searching and information retrieval. In: *Journal of the American Society for Information Science and Technology*, 61(8), 1517–1534.
- [8] Brandstetter, Th.; Hübel, Th.; Tantner, A. (Hrsg.) (2012): *Vor Google. Eine Mediengeschichte der Suchmaschine im analogen Zeitalter*. Bielefeld: transcript.
- [9] Head, A.; Eisenberg, M. (2009): *How College Students Seek Information in the Digital Age*. University of Washington. (Project Information Literacy Progress Report)
- [10] Connaway, L.S.; Dickey, T.J. (2010): *The digital information seeker: Findings from selected OCLC, RIN and JISC user behaviour projects*. London: JISC.
- [11] JISC (2010): *User Behaviour Observational Study. User Behaviour in Resource Discovery*. London: JISC.
- [12] JISC (2008): *Information behaviour of the researcher of the future*. London: JISC. (Teilstudien: *Trends in Scholarly Information Behaviour; The Literature on Young People and their Information Behaviour; Age-Related Information Behaviour; Student Information-Seeking Behaviour in Context; Technology Trends*)
- [13] OCLC (2008): *Online Catalogs: What Users and Librarians Want*. [Dublin, Ohio: OCLC]
- [14] Rowlands, I.; Nicholas, D.; Williams, P.; Huntington, P.; Fieldhouse, M.; Gunter, B.; Withey, R.; Jamali, H.; Dobrowolski, T.; Tenopir, C. (2008): The Google generation: the information behaviour of the researcher of the future. In *Aslib Proceedings*, 60(4), 290–310.
- [15] Buchanan, G.; Cunningham, S.; Blandford, A.; Rimmer, J.; Warwick, C. (2005): Information seeking by humanities scholars. In: *Research and Advanced Technology for Digital Libraries*, 218–229.
- [16] Frants, V.; Shapiro, J.; Voiskunskii, V. (1997): *Automated Information Retrieval. Theory and Methods*. San Diego u.a.: Academic Press.
- [17] Stock, W. G. (2007): *Information Retrieval. Informationen suchen und finden*. München: Oldenbourg.
- [18] Broder, A. (2002): A taxonomy of web search. In: *ACM Sigir forum*, 36(2), 3–10.
- [19] Marchionini, G. (2006): Exploratory search: from finding to understanding. In: *Communications of the ACM*, 49(4), 41–46.
- [20] Wildemuth, B. (2003): The effects of domain knowledge on search tactic formulation. In: *Journal of the American Society for Information Science and Technology* 55(3), 246–258.
- [21] White, R.; Dumais, S.; Teevan, J. (2009): Characterizing the influence of domain expertise on web search behavior. In: *Proceedings of the Second ACM International Conference on Web Search and Data Mining. Barcelona, Spain*. New York: ACM, 132–141.
- [22] Singer, G.; Norbirsath, U.; Vainikko, E.; Kikkas, H.; Lewandowski, D. (2011): Search-Logger Analyzing Exploratory Search Tasks. In: *SAC '11 Proceedings of the 2011 ACM Symposium on Applied Computing*. Taichung, Taiwan. New York: ACM, 751–756.
- [23] White, R. W.; Muresan, G.; Marchionini, G. (2006): *EESS 2006: Evaluating exploratory search systems. Proceedings of the ACM SIGIR 2006 Workshop*. [Redmond: Microsoft]
- [24] White, R. W.; Kules, B.; Drucker, S. M.; schraefel, m.c. (2006): Supporting Exploratory Search, Introduction, Special Issue *Communications of the ACM*, 49(4), 36–39.
- [25] Bates, M. (1989): The design of browsing and berrypicking techniques for the online search interface. In: *Online Review*, 13(5), 407–424.

- [26] Töpfer, A. (2008): *Erfolgreich Forschen: Ein Leitfaden für Bachelor-, Master-Studierende und Doktoranden*. Berlin, Heidelberg: Springer.
- [27] Corsten, H.; Deppe, J. (2008): *Technik des wissenschaftlichen Arbeitens*. 3. Aufl. München: Oldenbourg.
- [28] Kornmeier, M. (2007): *Wissenschaftstheorie und wissenschaftliches Arbeiten: eine Einführung für Wirtschaftswissenschaftler*. Heidelberg: Physica.
- [29] Merton, R.; Barber, E. (2004): *The travels and adventures of serendipity: A study in sociological semantics and the sociology of science*. Princeton, NJ: Princeton University Press.
- [30] de Rond, M.; Morley, I. (eds.) (2010): *Serendipity: fortune and the prepared mind*. Cambridge, New York: Cambridge University Press.
- [31] Reynolds, D.; Dickinson, I.; Grosnevov, D. (2009): *A query refinement model for exploratory semantic search*. Technical Report HPL-2009-167. [Palo Alto:] HP Laboratories.
- [32] Kules, B.; Capra, R. (2009): Designing Exploratory Search Tasks for User Studies of Information Seeking Support Systems. In: *JCDL '09 Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, 419–420.
- [33] Elzer, P.; Krohn, U. (1997): Visualisierung zur Unterstützung der Suche in komplexen Datenbeständen. In: *Proceedings of HIM: Hypertext – Information Retrieval – Multimedia '97: Theorien, Modelle und Implementierungen integrierter elektronischer Informationssysteme*. Universitätsverlag Konstanz, 27–38.
- [34] Glowacka, D.; Konyushkova, K.; Ruotsalo, T.; Kaski, S. (2012): Towards Exploratory Search of Scientific Information. [ECML-PKDD 2012, Paper retrieved at [http://adrem.ua.ac.be/iid2012/papers/glowacka\\_et\\_al-exploratory\\_search\\_of\\_scientific\\_information.pdf](http://adrem.ua.ac.be/iid2012/papers/glowacka_et_al-exploratory_search_of_scientific_information.pdf)]
- [35] Waitelonis, J.; Knuth, M.; Wolf, L.; Hercher, J.; Sack, H. (2010): The Path is the Destination – Enabling a New Search Paradigm with Linked Data. In: *LDFI-2010: Proceedings of the Workshop on Linked Data in the Future Internet at the Future Internet Assembly*. CEUR Workshop Proceedings, Vol. 700.
- [36] Wilson, M.; Kules, B.; Shneiderman, B. (2010): From keyword search to exploration: Designing future search interfaces for the Web. In: *Foundations and Trends in Web Science*, 2(1), 1–97.
- [37] Plieninger, J. (2011): Tools für das Monitoring. In: Lewandowski, D. (Hrsg.) (2011): *Handbuch Internetsuchmaschinen 2. Neue Entwicklungen in der Web-Suche*. Heidelberg: Aka, 323–177.
- [38] Saracevic, T. (2007): Relevance: A Review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. In: *Journal of the American Society for Information Science and Technology*, 58(13), 1915–1933.
- [39] Hjørland, B. (2010): The foundation of the concept of relevance. In: *Journal of the American Society for Information Science and Technology*, 61(2), 217–237.
- [40] Rieh, S.; Belkin, N. (2000): Interaction on the Web: Scholars' Judgment of Information Quality and Cognitive Authority. In: *Proceedings of the 63rd Annual Meeting of the ASIS*. Medford, NJ: Information Today for ASIS, 25–28.
- [41] Furnas, G.W.; Landauer, T.K.; Gomez, L.M.; Dumais, S.T. (1987): The vocabulary problem in human-system communication. In: *Communications of the ACM*, 30(11), 964–971.
- [42] Strohmaier, M.; Kröll, M.; Körner, Ch. (2009): Intentional query suggestion: making user goals more explicit during search. In: *Proceedings of the 2009 workshop on Web Search Click Data WSCD '09*. Barcelona, Spain. New York: ACM, 68–74.
- [43] Hearst, M. A. (2009): *Search user interfaces*. New York: Cambridge University Press.
- [44] Mayr, P.; Petras, V. (2008): Cross-concordances: terminology mapping and its effectiveness for information retrieval. In: *74th IFLA World Library and Information Congress*. Québec, Canada: IFLA.
- [45] Hienert, D.; Schaer, P.; Schaible, J.; Mayr, P. (2011): A Novel Combined Term Suggestion Service for Domain-Specific Digital Libraries. In: *International Conference on Theory and Practice of Digital Libraries (TPDL)*. Berlin: Springer, 192–203.
- [46] Lüke, T.; Schaer, P.; Mayr, P. (2012): Improving Retrieval Results with discipline-specific Query Expansion. In: *International Conference on Theory and Practice of Digital Libraries (TPDL 2012)*. Paphos, Cyprus. Berlin, Heidelberg: Springer, 408–413.
- [47] Mutschke, P.; Mayr, P.; Schaer, P.; Sure, Y. (2011): Science models as value-added services for scholarly information systems. *Scientometrics*, 89(1), 349–364.
- [48] Mayr, P. (2010): *Information Retrieval-Mehrwertdienste für Digitale Bibliotheken: Crosskonkordanzen und Bradfordizing* (GESIS-Schr.). Bonn: GESIS.
- [49] Hearst, M. A. (2000): Next generation web search: Setting our sites. In: *IEEE Data Engineering Bulletin*, 23(3), 38–48.
- [50] Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., Yee, K.-P. (2002): Finding the flow in web site search. In: *Communications of the ACM*, 45(9), 42–49.

- [51] Tombros, A., & Sanderson, M. (1998). Advantages of query biased summaries in information retrieval. In: *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* New York, NY, USA: ACM. doi:10.1145/290941.290947 2-10.
- [52] Quirnbach, S. (2009): Universal search: Kontextuelle Einbindung von Ergebnissen unterschiedlicher Quellen und Auswirkungen auf das User Interface. In: Lewandowski, D. (2009) (Hrsg.): *Handbuch Internet-Suchmaschinen: Nutzerorientierung in Wissenschaft und Praxis*. Heidelberg: Aka, 220–248.
- [53] Lalmas, M. (2011): Aggregated Search. In: M. Melucci; R. Baeza-Yates (eds.): *Advanced Topics in Information Retrieval*. Berlin, Heidelberg: Springer, 109–123.
- [54] Arguello, J., Diaz, F., Callan, J. (2011): Learning to aggregate vertical results into web search results. In: *Proceedings of the 20th ACM international conference on Information and knowledge management*, 201–210.
- [55] Lin, J.; Wilbur, W. J. (2007). PubMed related articles: a probabilistic topic-based model for content similarity. In: *BMC bioinformatics*, 8(1), 423.
- [56] Sawitzki, F.; Zens, M.; Mayr, P. (2013): Referenzen und Zitationen zur Unterstützung der Suche in SOWIPORT. In: *Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten. Proceedings des 13. Internationalen Symposiums der Informationswissenschaft (ISI 2013)*.
- [57] Schaer, P.; Mayr, P.; Lüke, T. (2012). Extending Term Suggestion with Author Names. In: *International Conference on Theory and Practice of Digital Libraries (TPDL 2012)*. Paphos, Cyprus. Berlin, Heidelberg: Springer, 317-322. doi:10.1007/978-3-642-33290-6\_34