



6. Suchmaschinenforschung in Anknüpfung an Wolf Stock: Ein Rückblick auf sieben Artikel aus Password 1999-2003

Erschienen als:

Lewandowski, Dirk: Suchmaschinenforschung in Anknüpfung an Wolf Stock: Ein Rückblick auf sieben Artikel aus Password 1999-2003. In: Dorsch, I.; Fietkiewicz, K.; Ilhan, A.; Meschede, C.; Siebenlist, T.; (Hrsg.): Facets of Wolf Stock: Festschrift zu Ehren von Wolfgang G. Stock. Frankfurt am Main: DGI. S. 71-85.

Von **Dirk Lewandowski**

6.1 Einleitung

Wolf Stock hat sich vor allem Anfang der 2000er-Jahre in einer mehrteiligen Serie in der Zeitschrift Password mit der Funktionsweise und der Qualität von Internet-Suchwerkzeugen auseinandergesetzt (Stock & Stock, 2000b, 2000a, 2001a, 2001b; Stock, 1999, 2000). Vier dieser Artikel sind gemeinsam mit Mechtild Stock entstanden; insofern knüpft die in diesem Aufsatz beschriebene Forschung nicht nur an die Forschung von Wolf Stock, sondern auch an das gemeinsam mit seiner Frau geschaffene Werk an, welches natürlich ebenso eine Würdigung verdient. Dies betrifft in erster Linie die hier als Basisliteratur verwendeten Password-Artikel, aber auch spätere Werke, insbesondere das umfangreiche „Handbook of Information Science“ (Stock & Stock, 2013).

Wolf Stock hat sich früh mit dem Themenfeld Suchmaschinen beschäftigt. Zwar waren um die Jahrtausendwende Suchmaschinen nicht mehr vollkommen neu, allerdings war gerade erst die erste Phase ihrer Entwicklung beendet (Van Couvering, 2008), während noch nicht entschieden war, in welche Richtung die weitere Entwicklung gehen würde.

In diesem Aufsatz werden die Password-Artikel wiedergelesen und in ihrer Stellung zur damaligen wie zur heutigen Zeit diskutiert. Es zeigt sich, dass viele der damals angesprochenen Themen noch heute relevant sind, sich aber auch Besonderheiten zeigen, die nur aus dem damaligen Kontext erklärbar sind. Insofern stellt dieser Aufsatz auch eine Reise in die Vergangenheit, genauer in die Zeit Ende der 1990er, Anfang der 2000er-Jahre, dar und taucht in die damalige Welt der Suchmaschinen ein, in der noch so viel mehr als heute möglich schien.

Die sieben Aufsätze sind allesamt in der Zeitschrift Password erschienen, einem Praktiker-Newsletter. Dies mag zuerst einmal ungewöhnlich erscheinen, wo doch

Wissenschaftler ihre Aufsätze zuvorderst in wissenschaftlichen Zeitschriften publizieren sollen. Die Informationswissenschaft ist häufig dafür kritisiert worden, dass sie in ihrer Forschung zu wenig Bezug auf Praktikerfragen wirft und ihre Ergebnisse, so sie denn für die Praxis relevant sind, zu wenig an die Praktiker vermittelt (s. dazu beispielsweise die Kritik von Bredemeier (2010)). Auf der anderen Seite wird von der Wissenschaft unter dem Schlagwort Third Mission immer stärker gefordert, ihren Nutzen über den direkten Bezug der Wissenschaft hinaus zu zeigen. Unter diesem Gesichtspunkt ist die Password-Serie geradezu vorbildlich: Sie vermittelt auf wissenschaftlichem Niveau und gleichzeitig für Nicht-Wissenschaftler verständlich, wie die damals noch relativ neuen Suchmaschinen funktionieren und zeigt, welche Stärken und Schwächen diese Suchwerkzeuge haben. Dies alles geschieht immer unter Rückgriff auf die klassischen Themen der Informationswissenschaft, so dass der fachliche Fokus stets gewahrt bleibt.

Den Artikeln ist gemeinsam, dass sie eine Perspektive zwischen der reinen Anleitung („Wie kann ich besser suchen?“, damals beispielsweise in der populären „Suchfibel“ (Karzauninkat, 1999) behandelt) und der sehr spezialisierten Literatur aus Fachzeitschriften, Konferenzbänden und Patenten einnehmen. Sie bewegen sich damit zwischen dem, was die Stocks später als „Praxisdiät“ bzw. „Theoriediät“ bezeichnet haben (Stock & Stock, 2004, S. 1). Es geht also um die Vermittlung zwischen den beiden Polen: Einerseits geht es um die Fundierung der Recherche bzw. die Analyse von Suchfunktionen, andererseits um die Vermittlung von Grundlagenwissen in die Praxis.

Aus den Password-Aufsätzen ist später auch ein Buch geworden, in welchem die Aufsätze leicht aktualisiert und um zwei weitere Artikel ergänzt erschienen sind (Stock & Stock, 2004). Hier soll allerdings auf die Originale zurückgegriffen werden; zum einen, weil sie noch unmittelbarer den Zeitbezug herausstellen, zum anderen, weil die zwei neu hinzugekommenen Aufsätze sich nur mittelbar mit Suchmaschinen beschäftigen. In ihnen geht es vielmehr um konkrete Informationsanbieter, die ihre Informationen über das Web verkaufen. Die im Folgenden zu besprechenden Aufsätze bieten nicht nur einen Blick in die Vergangenheit, sondern sie zeigen auch eine Perspektive auf Suchmaschinen, die heute weitgehend fehlt, nämlich die der professionell Recherchierenden. Und, vielleicht noch wichtiger, zeigen diese Artikel ein Dilemma in der heutigen Betrachtung der Suchmaschinen: Wir fragen gar nicht mehr, wie die Suchwerkzeuge eigentlich sein könnten, sondern sind fixiert auf Google und nehmen an, dass Google hinsichtlich der Suche alles kann. Insofern zeigen die Artikel auch die Unterschiede zwischen einer Suchmaschinenlandschaft damals und einer weitgehenden Monokultur heute (Lewandowski, 2016b).

6.2 Die Suchmaschinenlandschaft 1999/2000

Um den Kontext der hier behandelten Aufsätze zu verstehen, ist es zunächst sinnvoll, einen Blick zurück auf die Suchmaschinenlandschaft um die Jahrtausendwende zu werfen. Leider gibt es keine Werke, die eine vollständige Geschichte der Suchmaschinen schreiben, sondern lediglich einige Beiträge, die entweder Teile der Suchmaschinengeschichte erzählen oder sich mit einzelnen Aspekten beschäftigen (u.a. Batelle (2005); Dominikowski (2013); van Couvering, 2008). Insofern kann auch an dieser Stelle nur ein schlaglichtartiger Blick auf die damalige Zeit geworfen werden; eine tiefergehende Beschäftigung mit der Suchmaschinengeschichte steht noch aus.

Zunächst einmal ist festzustellen, dass es in den betrachteten Jahren viele Dienste, die uns heute selbstverständlich sind, schlicht noch nicht gab. Dies betrifft unter anderem Wikipedia (gestartet 2001), Facebook (2004) und Twitter (2006). Generell gab es die heute selbstverständlichen sozialen Medien noch nicht; auf den engeren Suchkontext bezogen gab es weder die Universal Search noch den Knowledge Graph. Die Situation bei den Suchmaschinen zeichnete sich vielmehr durch viel Ausprobieren von neuen Funktionen aus; gerade dies macht den Blick zurück in diese Phase so spannend.

Ein wichtiger Punkt, der auch in den Password-Aufsätzen deutlich wird, ist die damalige Auswahl an Suchmaschinen. Während wir heute selbstverständlich davon ausgehen, dass es mit Google die eine große Suchmaschine gibt, die uns zu allen Informationsbedürfnissen weiterhelfen kann, haben Stock und Stock in ihrem Retrievaltest (Stock & Stock, 2000b) insgesamt 18 verschiedene Suchmaschinen behandelt; an keiner Stelle findet sich dort eine Aussage, dass dies eine besonders große Menge wäre oder bestimmte Suchmaschinen nur der Vollständigkeit halber hinzugenommen worden wären.

Wie stark die Fokussierung auf Google heute ist, zeigt exemplarisch die Ankündigung eines im Jahr 2019 veröffentlichten Suchmaschinentests auf der Titelseite der Computerzeitschrift c't: „Googeln ohne Google“ (Bager, 2019). Erfreulich ist jedoch, dass Alternativen zu Google nach längerer Zeit inzwischen wieder ins Bewusstsein zu rücken scheinen, wenn nun auch neben der Frage nach relevanten Ergebnissen und den Suchfunktionen häufig vor allem im Kontext eines besseren Datenschutzes.

Auch wenn die Auswahl der Suchmaschinen um die Jahrtausendwende erst einmal sehr groß war, so darf das nicht über die aus heutiger Sicht größtenteils katastrophal schlechten Suchergebnisse hinwegtäuschen. Beispielhaft kann hier die Untersuchung von Stock und Stock (Stock & Stock, 2000b) zu den Known-Item-Suchen angeführt werden, die in vielen Fällen zeigt, dass die damaligen Suchmaschinen nicht in der Lage waren, die einfache Suche nach Websites/Homepages zu erfüllen. Dass die Ergebnisse der Suchmaschinen damals von den Nutzern überhaupt akzeptiert wurden, lässt sich wohl nur dadurch erklären, dass man immer noch fasziniert davon war, was überhaupt alles mit einer einfachen und kostenlosen Suche zu finden war. Es war damals keineswegs selbstverständlich, zu allen möglichen Themen etwas Relevantes finden zu können.

Weiterhin fällt auf, dass damals bei weitem noch nicht entschieden war, wie Suchmaschinen Geld verdienen sollten. Auch in den Password-Aufsätzen wird dieses Thema immer wieder angesprochen; genannt werden unter anderem der Weg einiger Suchmaschinen, sich zu Portalen zu entwickeln sowie die Werbefinanzierung, die sich in die heute gängigen kontextbasierten Textanzeigen (Jansen, 2011; Lewandowski, 2018); und die heute in Suchmaschinen unübliche Bannerwerbung unterteilt.

Selbstverständlich standen um die Jahrtausendwende die beiden Ansätze der Erschließung des Web mittels algorithmischer Suchmaschinen und mittels Web-Verzeichnisse nebeneinander. So schreiben Stock und Stock (2000b): Die „Indexierung verläuft entweder automatisch via maschinelle Indexierung oder intellektuell durch die Einordnung von Dokumenten in Klassen.“ (S. 24) Zwar lässt sich in den Password-Aufsätzen schon die Tendenz erkennen, dass die Verzeichnisse durch die damalige Phase des sehr schnellen Wachstums des Web (dargestellt in Risvik und Michelsen (2002)) an Boden verloren; dass die Verzeichnisse aber weitestgehend

verschwinden würden, war noch nicht vorzusehen.

Besonders für den Kontext der professionellen Fachinformation, in dem ja die Password-Aufsätze veröffentlicht wurden, ist von Bedeutung, dass in den Jahren um die Jahrtausendwende noch nicht klar war, wie sich die beiden Bereiche „Kostenlosinformation im Web“ und kostenpflichtige Fachinformation zueinander verhalten sollten. Inzwischen ist seit langem klar, dass sich die Fachinformationsanbieter den Suchmaschinen öffnen mussten und auch geöffnet haben. Suchmaschinen werden als „Schaufenster“ der Dienste verwendet; mittels Suchmaschinenoptimierung werden Nutzer über die allgemeinen Suchmaschinen angelockt.

6.3 Die Password-Aufsätze

Im Folgenden werden nun die sieben Aufsätze einzeln besprochen. Dabei werde ich auf die aus meiner Sicht wesentlichen Punkte eingehen, ohne dass es sich um eine systematische Auswertung der dort vorgebrachten Fakten und Ansichten handelt.

6.3.1 Stock, W. G. (1999). RealNames: Verkaufte Suchwörter, verkaufte Links. Retrieval nach Homepages optimiert?

RealNames war ein Dienst, bei dem Unternehmen für Suchbegriffe die Anzeige ihrer Website als Ergebnis einer Suchanfrage kaufen konnten. Eingebunden wurde RealNames in bekannte Suchmaschinen wie AltaVista, auf deren Trefferseiten das Ergebnis von RealNames vor den eigentlichen Ergebnissen angezeigt wurde. Das System griff auf unterschiedliche Quellen zurück, um die Homepage-Suche zu ermöglichen. Stock (1999) nennt verkaufte Suchwörter, intellektuell erschlossene Homepages, die (automatische) Analyse des Domainnamens und die Kooperation mit anderen Informationsproduzenten. Damit steht bei diesem Dienst der Nutzen im Vordergrund und er unterscheidet sich durch die Kombination von bezahlten und unbezahlten Einträgen von den heutigen Textanzeigenblöcken, in denen nur erscheint, wer auch bezahlt hat.

RealNames erinnert an die heutigen Möglichkeiten der Anzeigenbuchung bei Suchmaschinen: Unternehmen bieten auf Suchbegriffe, um für diese innerhalb des Anzeigenblocks, der den organischen Suchergebnissen vorgeschaltet ist, bevorzugt gelistet zu werden. Allerdings begegnet RealNames im Gegensatz zur heutigen Anzeigenplatzierung einem damals gravierenden Problem, nämlich dass Suchmaschinen größtenteils nicht in der Lage waren, zu navigationsorientierten Suchanfragen (also der Suche nach einer Website, die bereits bekannt ist bzw. deren Existenz vom Suchenden zumindest vermutet wird) das jeweils richtige Ergebnis zu finden. Beispielhaft zeigt dies Stock anhand der Suche nach der Homepage von Bill Clinton. In der Suchmaschine AltaVista findet sich dessen Homepage auf der Website des Weißen Hauses nicht einmal unter den ersten 50 Treffern, eine aus heutiger Sicht unvorstellbare Situation. Stock kommentiert: „Ohne das Verfahren von RealNames ist mit einem – zugegebenermaßen primitiven – Suchargument wie ‚Bill Clinton‘ die erwünschte Homepage praktisch kaum zu finden.“ (Stock, 1999, S. 22) Er schätzt damit die Situation so ein, dass das Problem alleine durch das Relevanzranking der Suchmaschinen nicht gelöst werden kann, sondern es eines Zusatzdienstes bedarf, der dies in Ergänzung zum Standardranking leistet.

Um die damalige Situation zu verstehen, muss man allerdings bedenken, dass die seit langem gebräuchliche Unterteilung von Suchanfragen in informations-

orientiert, navigationsorientiert und transaktionsorientiert (Broder, 2002) damals noch nicht existierte und die Suchmaschinenbetreiber erst zu begreifen begannen, dass sie bei der „Homepage-Suche“ ein enormes Problem hatten. Man könnte nun meinen, dass das Problem inzwischen vollständig gelöst wurde. Wie in einer späteren Studie (Lewandowski, 2015) gezeigt werden konnte, ist dies jedoch nicht der Fall, wenn auch vor allem Google für die navigationsorientierten Suchanfragen im Schnitt sehr gute Ergebnisse liefert.

Interessant ist Stocks Bewertung von RealNames als kommerziellem Dienst. Er schreibt: „Machen wir uns nichts vor: Das World Wide Web ist ein kommerzielles Unternehmen, die Dienstleistungen müssen finanziert werden. Und warum soll nicht ein Privatunternehmen mit Inhaltserschließung Geld verdienen?“ (Stock, 1999, S. 21) Während wir heute das Web ganz selbstverständlich als kommerzialisiert betrachten, gab es Ende der 90er-Jahre noch die starke Tendenz, das Web gegen eine Kommerzialisierung zu verteidigen. Wolf Stock hat sich schon damals klar auf die wirtschaftliche Seite gestellt und entsprechend auch Unternehmen empfohlen, sich mittels RealNames in den Suchmaschinen zu platzieren: „Eigentlich sollten alle Eigennamen (wie z.B. Unternehmens-, Produkt- und Personennamen), als Suchargument eingegeben, zur Homepage des Namensträgers führen.“ (Stock, 1999, S. 23) Während dies heute kein Problem mehr darstellt, hat sich das Problem in gewisser Weise nur verschoben: Da die Suchmaschinennutzer zu einem großen Teil nicht zwischen den Werbetreffern und den organischen Suchergebnissen unterscheiden können (Lewandowski, Kerkmann, Rümmele & Sünkler, 2018), reicht es für Unternehmen nicht aus, für ihren Namen als Suchbegriff auf Platz 1 der organischen Treffer zu stehen. Vielmehr sehen sie sich häufig gezwungen, auf ihren eigenen Unternehmens- oder Markennamen zu bieten, um auch im Anzeigenblock, der von vielen Nutzern nicht als solcher erkannt wird, auf dem ersten Rang angezeigt zu werden.

Der Aufsatz zu RealNames zeigt also zwei miteinander verbundene Themen auf, die auch heute noch eine wichtige Rolle spielen, nämlich die Vermarktung von Suchbegriffen und die Bearbeitung von navigationsorientierten Suchanfragen. Dabei muss man sich noch einmal vergegenwärtigen, dass die Positionen in der damaligen Zeit zwischen den Extremen „bezahlte Treffer bringen bessere Suchergebnisse als das Ranking“ (vertreten u.a. in einem von GoTo angemeldeten Patent; Davis et al., 2001) bis zu „Werbung hat einen negativen Einfluss und hat in Suchmaschinen nichts zu suchen“ (vertreten ausgerechnet von den Google-Gründern Larry Page und Sergej Brin; Brin & Page, 1998) lagen. Heute ist das Pendel eindeutig in Richtung der Werbung ausgeschlagen. Auch wenn zumindest eine Kennzeichnung der durch Bezahlung entstandenen Treffer vorliegt, steht deren Erkennbarkeit bzw. tatsächliche Erkennung durch die Nutzer immer wieder in der Kritik (Sullivan, 2013).

6.3.2 Stock, W. G. (2000). Checkliste für Retrievalsysteme: Qualitätskriterien von Suchmaschinen

In der von Wolf Stock (2000) entwickelten Checkliste für Retrievalsysteme geht es nicht nur, aber auch um Suchmaschinen. Stock widmet sich in diesem Aufsatz der Frage, welche Funktionen ein solches System zur Verfügung stellen sollte, um allen Anforderungen auch des professionellen Rechercheurs gerecht werden zu können. Der hohe Anspruch zeigt sich dabei unter anderem in folgendem Zitat: „Wichtig für den Nutzer ist ein reibungsloses Zusammenspiel von Basisfunktionen,

informatrischen Befehlen und natürlichsprachigen Aspekten, so etwa das Weiterbearbeiten einer im mengentheoretischen Retrieval gefundenen Treffermenge mittels Relevance Ranking oder die Boolesche Bearbeitung einer natürlichsprachig erhaltenen Treffermenge.“ (Stock, 2000, S. 31) Der entstandene Kriterienkatalog gliedert sich entsprechend in die Bereiche Basisfunktionen, informatrische Funktionen (zur Beschreibung der Treffermengen) und Kriterien zur Mächtigkeit automatischer Indexierung und natürlichsprachiger Suche.

Stock sieht seine Checkliste als „eine Minimalforderung an Retrievalsysteme, die durchaus erweiterungsfähig ist.“ (Stock, 2000, S. 22) Dabei ist klar, dass, wenn man dem Autor darin folgen mag, dass es sich tatsächlich um Minimalanforderungen handelt, dies nur die Minimalanforderungen im Sinne eines professionellen Rechercheurs sein können.

Betrachtet man heutige Suchmaschinen (und nur um diese soll es hier ja gehen), zeigt sich, dass diese von den Stockschen Minimalanforderungen weit entfernt sind. Insbesondere Google ist seit langem für eine nur beschränkte Menge von erweiterten Suchfunktionen, die sich je nach Spezialsuche erheblich unterscheiden können, bekannt. Erschwerend kommt hinzu, dass einige dieser erweiterten Suchfunktionen in der Praxis nur eingeschränkt funktionieren (Bradley, 2017). Erfreulich ist immerhin, dass beispielsweise Bing dem Profi-Rechercheur ein weit größeres Repertoire an Suchbefehlen zur Verfügung stellt (Lewandowski, 2018, S. 228).

Betrachtet man die Bücher, die Suchenden helfen sollen, auch die erweiterten Funktionen von Suchmaschinen zu nutzen (z.B. Bradley, 2017), so zeigt sich eine deutliche Stärke der Stockschen Liste: Sie bietet einen klaren Anhaltspunkt dafür, was eine gute Suchfunktionalität wäre, während die Literatur in aller Regel Suchfunktionen aufzählt und beschreibt, ohne sie in den Kontext des eigentlich Wünschenswerten bzw. Erforderlichen zu stellen. Insofern ist Stocks Liste auch ein Beitrag zur umfassenden Qualitätsbewertung von Suchmaschinen, wie ich sie zusammen mit Nadine Höchstötter vorgeschlagen habe (Lewandowski & Höchstötter, 2007). Damit genügt die Liste dem von Stock selbst formulierten Anspruch, dass die „entstehende Sammlung von Qualitätskriterien (...) bei der Evaluation konkreter Systeme eingesetzt werden können (soll).“ (Stock, 2000, S. 22) Allerdings muss die Liste natürlich auch im Kontext der Zeit gesehen werden, in der sie entstanden ist. Damals schien es noch durchaus plausibel, dass sich auch die Web-Suchmaschinen zu Suchwerkzeugen entwickeln könnten, die auch den Ansprüchen professioneller Rechercheure genügen würden. Anhaltspunkte dafür gaben beispielsweise AltaVista mit seinem starken Suchvokabular und NorthernLight mit seiner Verbindung aus Web-Suchmaschine und kostenpflichtigen Datenbanken. Ein erneutes Lesen der Checkliste zeigt allerdings auch, welche Möglichkeiten zur differenzierten Recherche es eigentlich gibt, und wie wenig davon in den heutigen Suchmaschinen umgesetzt ist. Natürlich sind die Suchmaschinen einen anderen Weg gegangen, nämlich den, das Interesse der Nutzer aus sehr kurzen Suchanfragen durch die Anreicherung mit Kontextinformationen zu antizipieren. Während also der Aufwand auf der Seite der Nutzer erheblich reduziert wurde, geht dies zu Lasten der Kontrolle durch den Nutzer.

6.3.3 Stock, M., & Stock, W. G. (2000a). Internet-Suchwerkzeuge im Vergleich. Teil 1: Retrievaltest mit Known Item Searches

Eine zentrale Frage, die sich durch alle Password-Artikel zieht, lautet: „Können Laien und Information Professionals im WWW gezielt suchen und finden?“ (Stock & Stock,

2000b, S. 23) Es geht also immer um beide Zielgruppen, und man spürt den Wunsch, dass sich die Suchmaschinen, die zur damaligen Zeit von Information Professionals heftig kritisiert (z.B. Allen, 2000) oder gar ignoriert wurden, so entwickeln sollten, dass sie beiden Gruppen gerecht werden. Dass es anders kam, wissen wir heute. Nach Stock und Stock stehen die damaligen Suchmaschinen vor vier Problemen: „(1) Auffinden der Dokumente, (2) Indexieren der gefundenen Dokumente, (3) Aktualisierung der Datenbasis, (4) Angebot eines Suchsystems“ (Stock & Stock, 2000b, S. 24). Der in dem vorliegenden Artikel beschriebene Test bezieht sich auf den letzten Bereich, wobei natürlich das Auffinden und Indexieren der Dokumente (also der Index) immer auch eine Rolle spielen, damit Dokumente überhaupt gefunden werden können.

Stock und Stock beziehen sich auf Suchanfragen aus unterschiedlichen Bereichen und kritisieren, dass sich die bisherigen informationswissenschaftlichen Retrievaltests auf den Informationsbedarf von Wissenschaftlern bezogen und damit den Laienanwender außen vor lassen würden. Diese Situation hat sich dann allerdings recht schnell geändert, und auch relativ bald darauf folgende Suchmaschinen-Retrievaltests haben Laienanfragen berücksichtigt (z.B. Griesbaum, Rittberger, & Bekavac, 2002).

Stock und Stock untersuchen in ihrem im Oktober 2000 durchgeführten Test 18 (!) Suchsysteme, was noch einmal die damals unübersichtliche Situation auf dem Suchmaschinenmarkt verdeutlicht. Im Ergebnis zeigt sich, dass Google mit 65% der Known Items die größte Availability der untersuchten Systeme vorweisen kann. Dieser Wert bezieht sich allerdings auf die Suche mit dem erweiterten Suchformular, die in dieser Studie als selbstverständlich mitbehandelt wird, was heute wohl undenkbar wäre.

Aus dem Ergebnis schließen Stock und Stock: „Es liegt demnach nahe, Suchwerkzeuge in Kombination abzufragen. Diesen Aspekt haben auch die Betreiber der Suchsysteme bemerkt und setzen auf Kooperation. Wir erinnern hier nur an die Zusammenarbeit von Yahoo! und Google sowie an Oingos Kooperationen mit AltaVista und dem Open Directory Projekt“ (Stock & Stock, 2000b, S. 30). Diese Empfehlung erscheint mir besonders interessant aus dem Grund, weil hier (wie an anderen Stellen der Artikel auch) keine Hoffnung auf eine grundsätzliche Verbesserung der Suchmaschinen, wie sie ja in der Realität später eingetroffen ist, zu erkennen ist. Vielmehr geht die Empfehlung dahin, mehrere Werkzeuge, die alleine jeweils nur unbefriedigende Ergebnisse erbringen, zu kombinieren. Diese Empfehlung ist sicher nicht falsch und besitzt für Recherchen mit dem Anspruch an Vollständigkeit nach wie vor Gültigkeit; das Known-Item-Problem ist aber schon seit längerem auch innerhalb einzelner Suchmaschinen weitestgehend gelöst. Nichtsdestotrotz sind Known-Item-Studien (bzw. Studien zu navigationsorientierten Suchanfragen) nach wie vor relevant (s. u.a. Behnert & Lewandowski, 2017; Lewandowski, 2011; Lewandowski, 2015).

6.3.4 Stock, M., & Stock, W. G. (2000b). Internet-Suchwerkzeuge im Vergleich (II): Klassifikation und terminologische Kontrolle: Yahoo!, Open Directory und Oingo im Vergleich.

Der dritte Artikel widmet sich den damals noch wichtigen Web-Verzeichnissen, also dem Ansatz, das Web „von Hand“ zu erschließen und zu klassifizieren. Die Artikelseerie beschäftigt sich damit neben der Beschreibung einzelner Suchmaschinen auch mit den grundsätzlichen Ansätzen zur Durchsuchbarmachung des Web, und hier

ist zuerst die damals noch gängige Unterscheidung zwischen (algorithmischen) Suchmaschinen und den manuell gepflegten Web-Verzeichnissen zu nennen.

Während Web-Verzeichnisse heute keine Rolle mehr spielen und den Kampf gegen die Suchmaschinen längst verloren haben (Lewandowski, 2018, S. 16ff.), waren die Verzeichnisse vor allem in der Anfangszeit des World Wide Web ein konkurrenzfähiger Ansatz. Einige Jahre später, also in der uns interessierenden Zeit der Entwicklung, waren sie schon ein Stück weit ins Hintertreffen geraten, vor allem durch das wachsende Web und die sich daraus ergebende Schwierigkeit, mit diesem Wachstum mitzuhalten, sowie durch die wachsenden Datenbestände der Suchmaschinen, die nun wesentlich mehr Inhalte auffindbar machten. Die Richtung der weiteren Entwicklung mag sich damals schon abgezeichnet haben, aber auch die Betreiber der Suchmaschinen setzten zumindest auf die Integration der Verzeichnisse in ihre Systeme. Und diese ging in beide Richtungen: Verzeichnisse wie Yahoo ergänzten ihre Dienste um Ergebnisse aus Suchmaschinen, während diese ihre Ergebnisse durch Verzeichnisse ergänzten. Der Artikel beschäftigt sich nun auf der einen Seite mit den von Yahoo und Open Directory eingesetzten Klassifikationen, auf der anderen Seite mit der Auffindbarkeit bzw. Suche innerhalb dieser Verzeichnisse.

Die Klassifikationen von Yahoo und Open Directory werden heftig – und zu Recht! – kritisiert. Anstatt sich an bestehenden Klassifikationen (und der Klassifikationstheorie) zu orientieren, würden die Web-Verzeichnisse das Rad neu erfinden, was ihnen und den Suchenden allerhand praktische Probleme verursachen würde. Hier nur zwei Beispiele:

- „Das Klassifikationsgerüst von Yahoo! ist – für eine Klassifikation überraschend an einigen Stellen polyhierarchisch organisiert. Die Klasse ‚Bibliotheks- und Informationswissenschaft‘ hat sowohl ‚Geisteswissenschaften‘ als auch ‚Bibliotheken‘ als Oberbegriffe.“ (Stock & Stock, 2000a, S. 29)
- „Die Einführung einer Klassifikationsebene, die nach dem Alphabet sortiert (im Beispiel ‚K‘), was beim ODP öfters vorkommt, kann wohl nur als Kapitulation vor den Problemen einer thematischen Ordnung gedeutet werden.“ (Stock & Stock, 2000a, S. 30)

Entsprechend fällt das Resümee zu den beiden Verzeichnissen aus. Zu Yahoo schreiben Stock und Stock: „Das leitende Ordnungskriterium bei Yahoo! ist uns verborgen geblieben. Wo ist der rote Faden? (...) Für das Herausgeberteam von Yahoo! besteht die Gefahr, dass ihnen die Systemordnung aus den Fingern gleitet.“ (Stock & Stock, 2000a, S. 29) Und zum Open Directory: „Die Probleme einer riesigen, überwuchernden präkombinierten Klassifikation bleiben jedoch erhalten.“ (Stock & Stock, 2000a, S. 30) Als Lösung schlagen Stock und Stock vor, dass sich Web-Verzeichnisse bestehende Klassifikationen die die DDC zur Grundlage nehmen sollten, die natürlich entsprechend angepasst werden müssten. Sie führen Beispiele für solche Projekte an, die sich jedoch in der späteren Praxis nicht durchgesetzt haben. Neben den im Artikel angeführten Beispielen ist hier sicher auch an die Suchmaschine GERHARD der Universitätsbibliothek Oldenburg zu denken (Wätjen, 1998).

Mit Blick auf die Suche stellt sich natürlich die Frage, wie die Verzeichnisse angesichts ihres relativ geringen Umfangs bzw. ihrer schlechten Abdeckung in Hinblick der Größe des Web einen positiven Beitrag zur Suche in dessen Datenbeständen leisten können. Stock und Stock sehen die Lösung in einer Kombination von Suchmaschinen und Web-Verzeichnissen, einem damals von den meisten

Anbietern verfolgten Ansatz. Sie beschreiben, wie beispielsweise in der Suche die Websites nicht nur in ihren Klassen gelistet, sondern wie sie innerhalb der Klassen als Antwort auf eine Suchanfrage sortiert werden können, wie ganze Klassen als Treffer in das Suchergebnis eingebunden werden können und wie (noch recht statisch) Verzeichnis und Suchmaschinenergebnisse auf einer Trefferseite kombiniert werden können. Die Kombination von Suchmaschine und Verzeichnis wurde von allen großen Anbietern noch einige Jahre weitergeführt; Google gab beispielsweise die Integration des Open Directory erst im Jahr 2011 endgültig auf.

Man muss sich aus heutiger Sicht klarmachen, dass die Idee der Klassifikation des Web kein Unsinn, sondern Standard war, und auch bei „normalen Suchmaschinen“ zusätzlich eingesetzt wurde (s. dazu auch Hamdorf (2001, 2004)).

Im Kontext der klassifikatorischen Erschließung des Web – aber nicht nur dort – stellt sich die Frage nach der Rolle von Bibliothekaren und Information Professionals. Stock und Stock zeigen hier anhand der Verzeichnisse, aber auch anhand der Terminologiedatenbank Oingo, dass das „traditionelle“ Know-How dringend gebraucht werden würde. Jürgen Krause hat es ein paar Jahre später so formuliert: „Bibliothekarische und informationswissenschaftliche Traditionen und Erkenntnisse spielen bei deren (der Suchmaschinen) Lösungsansätzen kaum eine Rolle. Das wäre positiv, wenn die neuen Ideen bessere Ergebnisse liefern, und schlecht, wenn es nur darauf beruht, dass der State-of-the-Art von den neuen Playern einfach nicht beachtet wird, weil die das mit dem Web sowieso nicht können“ (Krause, 2004). Stock und Stock zeigen in ihrem Aufsatz bereits, wo Schwierigkeiten entstehen können, wenn man versucht, das Rad neu zu erfinden anstatt auf bestehendem Wissen aufzubauen.

6.3.5 Stock, M., & Stock, W. G. (2001). Internet-Suchwerkzeuge im Vergleich (III) Informationslinguistik und -statistik. AltaVista, FAST und Northern Light.

In diesem Artikel (Stock & Stock, 2001a) werden drei Suchmaschinen beschrieben, die recht unterschiedliche Ansätze verfolgen. Während es sich bei AltaVista um den damaligen Standard einer Suchmaschine handelt, geht es bei der Beschreibung von FAST vor allem um deren Möglichkeiten der Suche nach multimedialen Inhalten (was damals bei weitem noch keine Selbstverständlichkeit war). Bei NorthernLight schließlich handelte es sich um eine Hybridsuchmaschine, also um ein System, das Inhalte aus dem Web mit Inhalten aus Fachdatenbanken kombinierte. Diesen Ansatz sehen Stock und Stock als besonders fruchtbar an: „Der von Northern Light eingeschlagene Weg der Hybride gehört u.E. weiterverfolgt und ausgebaut.“ (Stock & Stock, 2001a, S. 27) Dazu kam bei NorthernLight die automatische Clusteranalyse, die eine Besonderheit dieser Suchmaschine darstellte.

Interessant erscheint mir aus heutiger Sicht besonders, wie verschieden die besprochenen Systeme sind. Heute ist es ja so, dass auch die verschiedenen Suchmaschinen zumindest an der Oberfläche nicht nur sehr ähnlich, sondern eigentlich alle „wie Google“ sind.

Der Artikel beschreibt die Suchsysteme technisch recht tiefgehend; die Beschreibung basiert vor allem auf der Patentliteratur, aus der sich einiges Aufschlussreiches entnehmen lässt. Die Leistung hier liegt in der für das Fachpublikum verständlichen Beschreibung des technischen State of the Art der damaligen Suchmaschinen. Dass sich aus der Patentliteratur immer wieder interessante Erkenntnisse über Suchmaschinen entnehmen lassen, zeigten dann später unter anderem die Bücher von Stephen Arnold (Arnold, 2005, 2007).

6.3.6 Stock, M., & Stock, W. G. (2001). Internet-Suchwerkzeuge im Vergleich (IV): Relevance Ranking nach 'Popularität' von Webseiten.

Dieser Artikel (Stock & Stock, 2001b) setzt die Beschreibung der damals führenden Suchmaschinen fort, diesmal mit der 1998 gestarteten Suchmaschine Google. Zunächst einmal schätzen Stock und Stock sie in vielen Punkten positiv ein. Zuerst ist da einmal das aufgeräumte Design: „Google besticht durch das Angebot intuitiv verstehbarer Suchbildschirme“ (Stock & Stock, 2001b, S. 20). Erstaunlich aus heutiger Sicht ist, dass das damals betont werden musste. Vergleicht man Google zu dieser Zeit aber mit anderen Suchmaschinen, so ist der Unterschied auf einen Blick zu sehen. Die Password-Artikel enthalten durchgängig Screenshots, die das schön illustrieren. Weiterhin positiv fällt der große Index auf. Wir befinden uns mitten in den „Size Wars“ der Suchmaschinen (Sullivan, 2005), bei denen jede Suchmaschine die jeweils anderen mit der Größe ihres Datenbestands zu übertrumpfen suchte. Im Weiteren wird die Funktionsweise des PageRank erklärt, die inzwischen als bekannt vorausgesetzt werden kann. Kommen wir stattdessen zu den Suchfunktionen, die eine weit schlechtere Bewertung erhalten: „Wie sucht Google? Hier muss man zunächst notieren, was Google an erwarteter Suchfunktionalität nicht bringt. Nicht unterstützt werden: Jokerzeichen für Fragmentierungen, Groß- bzw. Kleinschreibung, Klammersetzung in der Suchfrage und automatische Wortstammbildung. Da nur exakt zeichengetreu gesucht wird, würden Tippfehler stets zu falschen Ergebnissen führen. Dieses Problem wird durch eine Tippfehleranalyse (bei Eingabe eines nicht bekannten Terms: „Did you mean: XXX?“) etwas gemildert. Voreingestellter Boolescher Operator ist UND. Disjunktive Suchen mit OR sind durchführbar, dabei bindet OR stärker als der Standardoperator UND“ (Stock & Stock, 2001b, S. 21).

Weiterhin beschäftigt sich der Artikel mit den bereits zahlreich in die Google-Ergebnisseiten eingebundenen Zusatzdiensten wie die bereits bekannten RealNames und Open Directory, aber auch Wörterbücher, Nachrichten von Fremdanbietern und Börsennotierungen. Dies kann als eine erste Übersicht zur weit später kommenden Universal Search (s. Quirnbach, 2009) gesehen werden.

6.3.7 Stock, W. G. (2003). Weltregionen des Internet: Digitale Informationen im WWW und via WWW

Der Aufsatz „Weltregionen des Internet: Digitale Informationen im WWW und via WWW“ (Stock, 2003) liefert gewissermaßen die Metaebene zu den bisher besprochenen Aufsätzen. In ihm wird eine Taxonomie der Online-Information beschrieben, welche diese Informationen grundlegend in die direkt im Web zugänglichen Informationen und die via Web zugänglichen Informationen einteilt. Die im Web vorhandenen Informationen sind zumindest grundsätzlich für Suchmaschinen (und auch Web-Verzeichnisse) zugänglich, auch wenn Inhalte aufgrund mangelnder Kapazität oder Technologie, aber auch aufgrund des willentlichen Ausschlusses von Inhalten in diesen Suchwerkzeugen fehlen. Gerade der willentliche Ausschluss ist m.E. ein Thema, das viel zu wenig Beachtung findet (s. Lewandowski, 2018, S. 44ff.). Stock schreibt dazu: „Teilweise ignorieren die Suchwerkzeuge aber auch „bewusst“ Material, und das findet ausdrücklich Zustimmung. Hier geht es um Spam, Informationsmüll, den keiner haben möchte und den auch keiner braucht - einige wenige Perverse und Waffennarren vielleicht ausgenommen. (Dass an dieser Stelle latent die Gefahr einer Zensur lauert, sei zugegeben.)“ (Stock, 2003, S. 26). Welche Inhalte nun wirklich „keiner braucht“, sei einmal dahingestellt. Richtig ist jedoch, dass Suchmaschinen, allein um ihre Funktionalität zu erhalten, Inhalte ausschließen

müssen. Gerade in aktuellen Diskussionen wie die um die „Upload-Filter“ könnte gerade dieser willentliche Ausschluss ein differenzierteres Bild vermitteln.

Auf der Seite der via Web zugänglichen Informationen stehen die Anbieter von Datenbanken, die über Web-Schnittstellen zugänglich sind, unabhängig davon, ob es sich um kostenlose oder kostenpflichtige Datenbanken handelt.

Aus meiner Sicht ist das Stocksche Modell für die Diskussion um das Invisible Web bzw. Deep Web eine wichtige Grundlage. Das Modell ermöglicht nicht nur eine Klassifikation von Suchdiensten, sondern durch die Einordnung von Informationsangeboten in dieses Modell könnten auch quantitative Verschiebungen zwischen den Bereichen aufgezeigt werden. Es ist offensichtlich, dass sich auf der Seite der Datenbanken eine Veränderung vollzogen hat, als dass deren Inhalte inzwischen fast selbstverständlich über Suchmaschinen zumindest in einer Vorschau verfügbar gemacht werden. Eine weitere quantitative Verschiebung zeigt sich bei der Betrachtung der Datenbestände einzelner Anbieter. Stock konnte 2001 noch schreiben: „Da allein ein Informationsanbieter wie Lexis-Nexis in etwa mit dem Suchmaschinen-Primus Google nach der Anzahl der Datensätze in etwa gleich zieht, ist es ganz sicher richtig, dass das Deep Web weitaus größer ist als das Oberflächenweb.“ (Stock, 2003, S. 27) Inzwischen hat das enorme Wachstum des Web und des Datenbestands von Google die Verhältnisse hier komplett verändert.

Eine auch nur halbwegs aktuelle Bestandsaufnahme des Deep Web gibt es leider nicht. Hier wäre das Modell von Stock ein hervorragender Ansatzpunkt. Leider wurde das Modell nie international publiziert, so dass es im nicht deutschsprachigen Bereich vielleicht einfach unbekannt ist. Die Literatur zum Deep Web (u.a. Devine & Egger-Sider, 2009, 2014) jedenfalls würde von einem systematischeren Ansatz erheblich profitieren. Stocks Schlussfolgerung aus dem Modell ist nun, dass für eine Vielzahl von Informationsbedürfnissen ein „Querweltein-Retrieval“ nötig wäre. Dem ist zuzustimmen, auch wenn durch die skizzierte Entwicklung der Öffnung der Datenbanken und die Einbindung zahlreicher Quellen direkt in die Suchmaschinen das Problem wohl eher aus dem allgemeinen Bewusstsein gerückt ist.

6.4 Was danach geschah

Den in den Password-Artikeln beschriebene Stand der Technik fassen Stock und Stock (Stock & Stock, 2001b) wie folgt zusammen: „Wir wollen ein vorläufiges Urteil wagen: Die Suchwerkzeuge sind primär auf den typischen Internetnutzer, der ohne dokumentarische oder andere informationspraktische Vorkenntnisse an seine Suche herangeht, zugeschnitten. Hier arbeiten sie derzeit gar nicht so schlecht. Information Professionals müssen sich notgedrungen auf diese Art des Retrieval einlassen“ (S. 27).

Dieser Befund gilt sicher heute noch genauso wie damals; man muss nur in die wenigen noch verbliebenen halbwegs aktuellen Recherchehandbücher (z.B. Bradley, 2017) schauen. Allerdings bleibt die Frage, ob die Suchmaschinen selbst jemals Werkzeuge für professionelle Rechercheure sein wollten. Aus meiner Sicht liegt hier eine zumindest damals populäre Fehleinschätzung vor, die unter anderem erklären könnte, warum nur so wenige Dienste, die selbst Web-Inhalte durchsuchbar machen, aus dem nicht-kommerziellen Umfeld entstanden sind. Wenn man annimmt, dass die großen Suchmaschinen ein Interesse daran haben, Dienste auch für die Profis zu entwickeln, dann sollte es ja nur eine Frage der Zeit sein, bis sie das auch tun.

Für Wolf Stock war das Thema Suchmaschinen nach der Password-Serie sicher nicht beendet, auch wenn er sich danach in seinen wissenschaftlichen Veröffentlichungen auf andere Themen verlegt hat. Vielleicht war es ja mehr das neue, das ihn an den Suchmaschinen interessiert hat – ebenso wie an den zahlreichen anderen neu aufgekommenen Themen, die er im Lauf der Jahre in seiner Forschung aufgenommen hat. Es ist aber natürlich schade, dass die Suchmaschinenforschung nicht systematisch in das Düsseldorfer Forschungsprogramm aufgenommen wurde, denn schon die in den Password-Artikeln besprochenen Themen und gesetzten Schwerpunkte zeigen einen lohnenswerten Weg auf.

Natürlich wurden Suchmaschinen in Wolf Stocks Werk immer wieder an verschiedenen Stellen thematisiert, dies allerdings vor allem zusammenfassend in einem größeren Kontext vor allem in den Lehrbüchern (Stock, 2007; Stock & Stock, 2013).

Die Arbeiten, die sicherlich am engsten an Stocks Arbeit zu Suchmaschinen anknüpfen, sind einige der in Düsseldorf entstandenen Dissertationen, die hier in chronologischer Reihenfolge kurz beschrieben werden sollen.

In meiner eigenen Dissertation (Lewandowski, 2005) habe ich Wolf Stocks Sichtweise insofern aufgenommen und weitergeführt, als dass die Arbeit einen umfassenden State-of-the-Art-Bericht liefert, an den sich konkrete Verbesserungsvorschläge für Bereiche anschließen, die für Information Professionals von besonderem Interesse sind. Das der Arbeit zumindest implizit zugrunde liegende Nutzermodell war also ebenso wie bei Stock das des Profi-Rechercheurs. Eine weitere Gemeinsamkeit ist die Einbindung der Websuche in die Historie von Suchsystemen und Erschließungsinstrumenten. Entsprechend schreibt Stock in seinem Vorwort zur Verlagsveröffentlichung der Dissertation: „Dabei zeigt sich, dass ein Bezug zu den „alten“ Erkenntnissen aus Informationswissenschaft und der Praxis der (ebenso ‚alten‘) professionellen Informationsdienste die aktuelle Diskussion befruchten kann“ (S. 10).

Die Arbeiten zur Taxonomie der Online-Information wurden von Matulová (2009) aufgenommen. Sie beschäftigt sich nicht nur mit dem Deep Web als solchem, sondern vor allem mit der Frage, wie Crosswalks (Übergänge) zwischen den Welten des Surface Web und des Deep Web geschaffen werden können.

Den Aspekt der Monetarisierung von Suchmaschinen griff Chang Kaiser (2010) in ihrer Arbeit zu Sponsored Links (Textanzeigen) auf. Sie beschreibt die beiden für Unternehmen interessanten Möglichkeiten der Suchmaschinenoptimierung und des Search Engine Advertising detailliert in ihren Aspekten und zeigt anhand einer Fallstudie, welche Besonderheiten auf dem chinesischen Markt gelten.

Bei der Arbeit von Sirotkin (Sirotkin, 2012) handelt es sich gewissermaßen um ein „Spiel über Bande“. Die Arbeit beschäftigt sich mit Kennzahlen für die Suchmaschinenevaluierung und knüpft damit vor allem an die Lehrveranstaltungen, die ich in Düsseldorf unter der Ägide von Wolf Stock gegeben habe, an. Sie verbindet das Thema Suchmaschinen mit grundlegenden Fragen des Information Retrieval und knüpft damit in diesem Bereich vor allem an den in Wolf Stocks Lehrbuch (2007) beschriebenen State of the Art an.

Tobias Siebenlist hat in seiner Dissertation (Siebenlist, 2015) eine Spezialsuchmaschine für emotional geladene Dokumente entwickelt und evaluiert. Damit erweitert er das Spektrum des Information Retrieval in Richtung emotionalem Information Retrieval und ergänzt das Methodenspektrum der Düsseldorfer Informationswissenschaft um die Softwareentwicklung.

Eine weitere Fortführung haben Wolf Stocks Arbeiten zu Suchmaschine natürlich

in der von mir geleiteten Forschungsgruppe Search Studies an der Hochschule für Angewandte Wissenschaften Hamburg gefunden. Auch wenn sich die Schwerpunkte in den mehr als zehn Jahren natürlich verändert haben, ist der Einfluss Wolf Stocks natürlich nach wie vor sichtbar.

6.5 Fazit

Die in diesem Aufsatz beschriebenen Artikel zeigen für den Bereich der Suchmaschinen, was damals war und was hätte sein können – oder sogar, was heute noch oder heute wieder gute Ideen wären. Ich möchte hier nur zwei Beispiele nennen:

- Die Verbindung von Web-Inhalten mit kostenpflichtigen Inhalten. Diese wird zugegebenermaßen durch die immer größeren Datenmengen immer schwieriger, auch wenn es mit dem Aufbau eines Offenen Web Index (Lewandowski, 2016a, 2019) m.E. eine naheliegende Basis für Lösungen gäbe.
- Die klassifikatorische Erschließung von Web-Inhalten, wofür heute ungleich bessere (automatisierte) Mittel zur Verfügung stehen als um die Jahrtausendwende.

Die Password-Artikel zeigen neben allen inhaltlichen Aspekten aber noch eine weitere große Stärke: Sie bieten fundierte Analysen zu den damals aktuellen Themen und stellen damit eine große Vermittlungsleistung in die Praxis dar. Es wäre mehr als wünschenswert, wenn sich auch heute Autoren finden ließen, die in dieser Tiefe aktuelle Trends aufgreifen und für die Zielgruppe der Informationspraktiker aufbereiten würden.

Literaturverzeichnis

- Allen, M. M. (2000). Six Degrees of Affiliation. *Online*, 24 (5), 49–51.
- Arnold, S. E. (2005). The Google Legacy: How Google's Internet Search is Transforming Application Software.
- Arnold, S. E. (2007). *Google Version 2.0: The Calculating Predator*. Infonortics.
- Batelle, J. (2005). *The Search: How Google and its Rivals Rewrote the Rules of Business and Transformed our Culture*. London: Portfolio. London (ua): Brealey.
- Behnert, C. & Lewandowski, D. (2017). Known-item Searches Resulting in Zero Hits: Considerations for Discovery Systems. *The Journal of Academic Librarianship*, 43 (2), 128–134.
- Bradley, P. (2017). *Expert Internet Searching*. Facet Publishing.
- Bredemeier, W. (2010). Kritik der Informationswissenschaft. Anmerkungen eines interessierten und besorgten Bürgers mit Common Sense. *Password* (07/08), 12–16.
- Brin, S. & Page, L. (1998). The Anatomy of a Large-scale Hypertextual Web Search Engine. *Computer networks and ISDN systems*, 30 (1-7), 107–117.
- Broder, A. (2002). A Taxonomy of Web Search. In *ACM Sigir forum* (Bd. 36, S. 3–10).
- Davis, D. J., Derer, M., Garcia, J., Greco, L., Kurt, T. E., Kwong, T., ... others (2001, Juli 31). *System and Method for Influencing a Position on a Search Result List Generated by a Computer Network Search Engine*. Google Patents. (US Patent 6,269,361)
- Devine, J. & Egger-Sieder, F. (2009). *Going Beyond Google: The Invisible Web in Learning and Teaching*. Facet London.

- Devine, J. & Egger-Sider, F. (2014). *Going Beyond Google Again: Strategies for Using and Teaching the Invisible Web*. Neal-Schuman Publishers.
- Dominikowski, T. (2013). Zur Geschichte der Websuchmaschinen in Deutschland. *Handbuch Internet-Suchmaschinen*, 3, 3–34.
- Griesbaum, J., Rittberger, M. & Bekavac, B. (2002). Deutsche Suchmaschinen im Vergleich: AltaVista. de, Fireball. de, Google. de und Lycos. de. In *ISI* (S. 201–224).
- Hamdorf, K. (2001). Wer katalogisiert das Web? Dokumentarische Arbeit als Big Business und Freiwilligen-Projekt. *Information Wissenschaft Und Praxis*, 52 (5), 263–270.
- Hamdorf, K. (2004). Jenseits von Google: Erschliessung und Recherche: von Internet-Angeboten durch Webkataloge. *Information Wissenschaft und Praxis*, 55 (4), 221–224.
- Jansen, J. (2011). *Understanding Sponsored Search: Core Elements of Keyword Advertising*. Cambridge University Press.
- Kaiser, C. (2010). *Suchmaschinenwerbung. Sponsored Links als Geschäftsmodell der Suchwerkzeuge*. Mit einer Fallstudie über chinesische Suchdienste. Hamburg: Kovač.
- Karzauninkat, S. (1999). *Die Suchfibel:(wie findet man Informationen im Internet?)*. Klett-Schulbuchverl.
- Krause, J. (2004). Zur Problematik eines Gedankenexperiments über die Informationsversorgung der Wissenschaften. In *Wissen in Aktion* (S. 167–180).
- Lewandowski, D. (2005). *Web Information Retrieval: Technologien zur Informationssuche im Internet*. DGI.
- Lewandowski, D. (2011). The Retrieval Effectiveness of Search Engines on Navigational Queries. In *Aslib Proceedings* (Bd. 63, S. 354–363).
- Lewandowski, D. (2015). Evaluating the Retrieval Effectiveness of Web Search Engines using a Representative Query Sample. *Journal of the Association for Information Science and Technology*, 66 (9), 1763–1775.
- Lewandowski, D. (2016a). Perspektiven eines Open Web Index. *Information-Wissenschaft & Praxis*, 67 (1), 15–21.
- Lewandowski, D. (2016b). Status Quo und Entwicklungsperspektiven des Suchmaschinenmarkts. *Handbuch Medienökonomie*, 1–23.
- Lewandowski, D. (2018). *Suchmaschinen verstehen*. Springer-Verlag.
- Lewandowski, D. (2019). The Web is Missing an Essential Part of Infrastructure: An Open Web Index. *arXiv preprint arXiv:1903.03846*.
- Lewandowski, D. & Höchstötter, N. (2007). Qualitätsmessung bei Suchmaschinen-System- und nutzerbezogene Evaluationsmaße. *Informatik-Spektrum*, 30 (3), 159–169.
- Lewandowski, D., Kerkmann, F., Rümmele, S. & Sünkler, S. (2018). An Empirical Investigation on Search Engine ad Disclosure. *Journal of the Association for Information Science and Technology*, 69 (3), 420–437.
- Matulová, H. N. (2009). *Crosswalks Between the Deep Web and the Surface Web*. Kovač.
- Quirnbach, S. (2009). Universal Search: Kontextuelle Einbindung von Ergebnissen unterschiedlicher Quellen und Auswirkungen auf das User Interface. In *Handbuch Internet-Suchmaschinen: Nutzerorientierung in Wissenschaft und Praxis* (S. 175–203). Akademische Verlagsgesellschaft AKA.
- Risvik, K. M. & Michelsen, R. (2002). Search Engines and Web Dynamics. *Computer*

- Networks*, 39 (3), 289–302.
- Sirotkin, P. (2012). On Search Engine Evaluation Metrics. *arXiv preprint arXiv:1302.2318*.
- Stock, M. & Stock, W. G. (2000a). Internet-Suchwerkzeuge im Vergleich (II): Klassifikation und terminologische Kontrolle: Yahoo!, Open Directory und Oingo im Vergleich. *Password* (12), 26–32.
- Stock, M. & Stock, W. G. (2000b). Internet-Suchwerkzeuge im Vergleich, Teil 1: Retrievaltest mit Known Item Searches. *Password*, 15 (11), 23–31.
- Stock, M. & Stock, W. G. (2001a). Internet-Suchwerkzeuge im Vergleich (III) Informationslinguistik und -statistik. AltaVista, FAST und Northern Light. *Password* (1), 16–24.
- Stock, M. & Stock, W. G. (2001b). Internet-Suchwerkzeuge im Vergleich (IV): Relevance Ranking nach „Popularität“ von Webseiten. *Password* (2), 20–27.
- Stock, M. & Stock, W. G. (2004). *Karrieretipp: Recherchieren im Internet* (Nr. 80). expert verlag.
- Stock, W. G. (1999). RealNames: Verkaufte Suchwörter, verkaufte Links. Retrieval nach Homepages optimiert? *Password* (11), 21–24.
- Stock, W. G. (2000). Checkliste für Retrievalsysteme: Qualitätskriterien von Suchmaschinen. *Password* (5), 22–31.
- Stock, W. G. (2003). Weltregionen des Internet: Digitale Informationen im WWW und via WWW. *Password* (2), 26–28.
- Stock, W. G. (2007). *Information Retrieval. Informationen suchen und finden*. München, Wien: Oldenbourg.
- Stock, W. G. & Stock, M. (2013). *Handbook of Information Science*. Walter de Gruyter.
- Sullivan, D. (2005). End of Size Wars? Google Says most Comprehensive but Drops Home Page Count. Retrieved May, 28, 2006.
- Sullivan, D. (2013). *Revealed: The 17 Other Search Engines the FTC Warned Over Paid Ad Disclosures*. Zugriff auf <http://searchengineand.com/revealed-17-other-search-engines-ftc-warned-168603>
- Van Couvering, E. (2008). The History of the Internet Search Engine: Navigational Media and the Traffic Commodity. In *Web search* (S. 177–206). Springer.
- Wätjen, H. (1998). GERHARD-automatisches Sammeln, Klassifizieren und Indexieren von wissenschaftlich relevanten Informationsressourcen im deutschen World Wide Web. *BIT online: Zeitschrift für Bibliothek, Information und Technologie*, 4, 279–290.