



Universität-Gesamthochschule Paderborn  
Fachbereich Informatik – FB 17  
Lehrstuhl für Betriebssysteme / Verteile Systeme

Seminararbeit

# **Suchmaschinen und ihr Nutzen für Electronic Commerce**

vorgelegt bei  
Prof. Dr. Hans-Ulrich Heiß

im Seminar  
Technologie für Electronic Commerce  
WS 1998 / 99

von

Carsten Schmidt

Studiengang Wirtschaftsinformatik

Matrikelnr. 3275954

## Inhaltsverzeichnis

1	Einleitung.....	3
1.1	Begriffsabgrenzungen und Entwicklungsgeschichte ...	3
1.2	Anforderungen an Suchmaschinen.....	4
1.3	Erfassungsbereich von Suchmaschinen.....	5
2	Funktionsweise von Suchmaschinen .....	5
2.1	Automatisches Indizieren.....	6
2.2	Extraktion von Metadaten .....	7
2.3	Suchanfrage.....	7
2.4	Verteilte und mehrfach verteilte Suchmaschinen .....	8
2.5	Metasuchmaschinen.....	8
3	Information Retrieval .....	10
3.1	Inverted Files .....	10
3.2	Stemming .....	11
3.3	Ranking / Weighting.....	11
3.4	Thesaurus.....	12
3.5	Stoplists.....	12
4	Fazit.....	13
4.1	Probleme existierender Suchhilfen .....	13
4.2	Nutzen für Electronic Commerce .....	14
4.3	Ausblick.....	15
	Literaturverzeichnis.....	16

### 1 Einleitung

Das World Wide Web stellt heute einen riesigen Wissensspeicher dar. Leistungsfähige Suchmaschinen werden benötigt, um in dieser heterogenen, weltweit verteilten Datenbank Informationen zu finden. Aufgrund der Tatsache, dass die meisten Inhalte in semistrukturierter Form vorliegen, versagen herkömmliche Abfragesprachen, die aus der Datenbanktechnologie bekannt sind.

In der vorliegenden Seminararbeit wird ein Überblick über die Entstehungsgeschichte der Suchmaschinen im World Wide Web und ihrer allgemeinen Funktionsweise gegeben. Darüber hinaus wird auf verteilte Suchmaschinen und Metasuchmaschinen eingegangen. Die wesentlichen Algorithmen und Datenstrukturen des Information Retrievals, deren sich Suchmaschinen bedienen werden vorgestellt. Abschließend wird auf Probleme, sowie den Nutzen für Electronic Commerce eingegangen und es wird ein Ausblick gegeben.

#### 1.1 Begriffsabgrenzungen und Entwicklungsgeschichte

In der noch recht jungen Geschichte des World Wide Web sind bereits eine Vielzahl von Suchhilfen entwickelt worden. Als erster Suchdienst ist hier WAIS (Wide Area Information Server) zu nennen, welcher um 1991 entwickelt wurde. Es handelt sich hierbei jedoch um einen lokale Indexierer, mit dem keine Suche über mehrere Server des World Wide Webs möglich ist. Ein weiterer Nachteil ist der große Bedarf an Speicherplatz für den Index, der in etwa dem der indizierten Dokumente entspricht. Dieses Manko wurde durch die Suchmaschine Glimpse 1993 ausgeglichen. Zeitgleich entstanden die ersten serverübergreifenden Suchmaschinen, wie Lycos, AltaVista, Yahoo, Excite oder Webcrawler<sup>1</sup>. Die sog. All-in-One-Formulare sind in der Lage, mehrere Suchdienste nacheinander über eine einheitliche Eingabemaske abzufragen. Hierbei entsteht allerdings kein Performancegewinn. Dieser ist erst durch den Einsatz von Metasuchmaschinen zu erreichen, die einfache Suchdienste parallel abfragen und

---

<sup>1</sup> Die Adressen lauten: <http://www.lycos.com>, <http://www.altavista.com>, <http://www.yahoo.com> und <http://webcrawler.com>

darüber hinaus noch weitere Vorteile aufweisen, worauf später detailliert eingegangen wird.

### **1.2 Anforderungen an Suchmaschinen**

Auf die Frage, welche Anforderungen an Suchmaschinen zu stellen sind, werden in [Mase97] und [Voss98] fünf zentrale Forderungen aufgestellt.

Von Suchdiensten wird verlangt, dass sie das gesamte World Wide Web durchsuchen. Dabei stellt sich die Frage, was unter dem gesamten World Wide Web zu verstehen ist. Handelt es sich hierbei um sämtliche im Internet verfügbaren Dokumente? Aufgrund der Größe des World Wide Webs und der Funktionsweise von Suchmaschinen, wie sie weiter unten erläutert wird, müssen hier Abstriche gemacht werden. In diese Problematik fällt auch die Suche nach dynamisch erzeugten Dokumenten, wie sie z. B. im Bereich des Electronic Commerce häufig zu finden sind.

Die gelieferten Suchergebnisse sollten insofern konsistent sein, indem sie kurzfristig nicht erreichbare oder ausgefallene Server berücksichtigen. Auch die dynamische Serverentwicklung (umbenennen von Verzeichnissen oder Dokumenten) muss beachtet werden.

Der Anwender soll neben der Möglichkeit, das Internet nach inhaltlichen Kriterien zu durchsuchen („Welche Dokumente behandelt ein bestimmtes Thema?“) auch in der Lage sein, eine strukturelle Anfrage an die Suchmaschine zu stellen („Welche Dokumente sind mit diesem durch Hyperlinks verknüpft?“).

Die Ergebnisse sollten für den Benutzer relevante Informationen enthalten. Es macht einen Unterschied, ob der gesuchte Begriff lediglich als Überschrift im Dokument enthalten ist, oder sich der Inhalt um genau dieses Thema dreht.

Schließlich sollte durch den Einsatz von Suchmaschinen kein Eingriff in die dezentrale und autarke Organisation des World Wide Web notwendig sein. Weder eine

Veränderung des HTTP-Protokolls<sup>2</sup>, noch zusätzliche HTML-Tags<sup>3</sup> sollen notwendig sein. Bei der großen Anzahl bestehender Dokumente ist eine komplette Überarbeitung ausgeschlossen.

Zusätzlich zu diesen fünf Anforderungen könnten auch schnelle Antwortzeiten, eine leichte Bedienbarkeit, sowie geringe Netzbelastung erfüllt werden, auf die allerdings im Rahmen dieser Arbeit nicht eingegangen wird.

### 1.3 Erfassungsbereich von Suchmaschinen

In [Lawr98] ist der Erfassungsbereich von Suchmaschinen untersucht worden. Hierbei stellten die Autoren fest, dass Suchmaschinen lediglich geringe Ausschnitte des World Wide Webs indizieren. HotBot lag hierbei mit 34% der Seiten an der Spitze, gefolgt von AltaVista (28%), Northern Light (20%), Excite (14%), Infoseek (10%) und Lycos mit 3%<sup>4</sup>. Der Einsatz aller sechs Suchmaschinen würde 60% der Dokumente im Internet abdecken.

## 2 Funktionsweise von Suchmaschinen

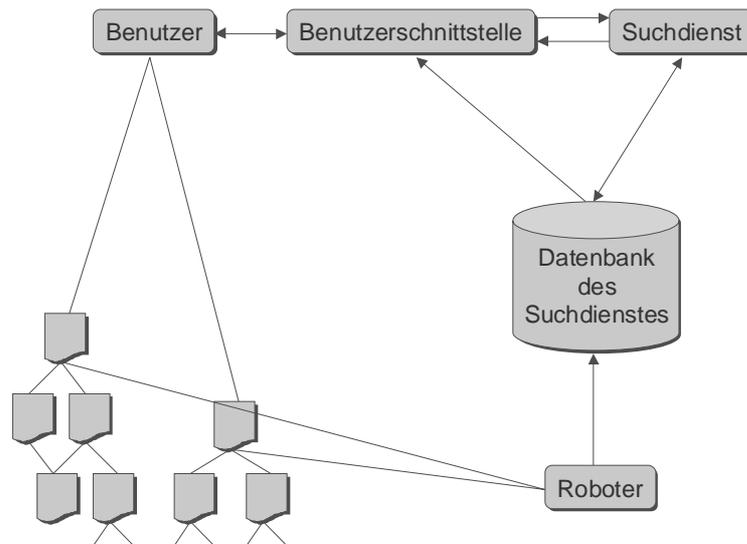
Eine Suchmaschine besteht im wesentlichen aus drei Teilen (s. Abbildung 1): 1) ein Metadaten sammelndes Programm, 2) den gesammelten Metadaten und 3) einer Schnittstelle, über die der Benutzer Anfragen an den Suchdienst stellt.

---

<sup>2</sup> HTTP (= Hypertext Transfer Protocol) bezeichnet das Protokoll, mit der HTML-Dokumente von einem Rechner auf den anderen übertragen werden.

<sup>3</sup> In HTML werden über Tags die Struktur sowie die Darstellung von Informationen festgelegt.

<sup>4</sup> Die Adressen lauten: HotBot (<http://www.hotbot.com>), Northern Light (<http://www.northernlight.com>), Excite (<http://www.excite.com>), Infoseek (<http://www.infoseek.com>)



**Abbildung 1: Ablauf einer Suchanfrage**

## 2.1 Automatisches Indizieren

Zunächst indiziert ein Roboter (auch Spider oder Worm genannt) die Dokumente des World Wide Webs. Ausgehend von der URL<sup>5</sup> eines Dokuments traversiert der Roboter das Internet in einem Breiten- oder Tiefendurchlauf. Beim Breitendurchlauf werden zunächst alle per Hyperlink referenzierten externen Dokumente abgesucht. Innerhalb dieser Dokumente wird genauso verfahren. Der Algorithmus stoppt nach einer vorher eingestellten Suchbreite oder sobald der gesamte Graph abgesucht ist.

Beim Tiefendurchlauf wird zuerst den Hypertextverweisen gefolgt, die auf einer tieferen Ebene des Verzeichnispfades liegen als das aktuelle Dokument. Dies geschieht solange, bis eine bestimmte Tiefe erreicht wurde oder der Algorithmus auf ein Dokument stößt, das keinen weiteren Hyperlink enthält. In diesem Fall wird einen Schritt zurückgegangen, um mit der nächsten Adresse fortzufahren.

---

<sup>5</sup> Uniform Resource Locator: Ein Uniform Resource Locator ist die Adresse eines Dokuments im WWW

### 2.2 Extraktion von Metadaten

Die auf diese Arten erreichten Dokumente werden einer lexikalischen Analyse unterzogen und die dabei gewonnenen Ergebnisse in die Metadatenbank des Suchdienstes eingestellt.

Bei dieser Extraktion der Metadaten setzen die Suchmaschinen unterschiedliche Verfahren ein. Zunächst bietet sich an, die gesamten Textinformationen innerhalb des Dokuments zu indizieren. Die hierbei angewendeten Verfahren des Information Retrievals werden im nächsten Kapitel ausführlich dargestellt. Dieses automatische Indizieren bringt jedoch Probleme mit sich. Kleinste Änderungen an den Dokumenten führen sofort zu einem Veralten der Metadaten. Auch ist so keine Gewichtung der Bedeutung der indizierten Daten möglich. Bedingt durch die Größe des Internets erfassen viele Suchmaschinen lediglich den Anfang des Dokuments.

Eine Alternative hierzu stellen die Metadaten des Dublin Core Element Sets [Dubl97] dar. Hierbei extrahiert ein Experte (meist der Autor) die Metadaten manuell aus dem Dokument. HTML bietet die Möglichkeit, die Metadaten ähnlich einem Kommentar dem Dokument hinzuzufügen. Das Dublin Core Element Set hat hierfür die 13 Schlüsselwörter *Title*, *Subject*, *Author*, *Publisher*, *OtherAgent*, *Date*, *ObjectType*, *Form*, *Identifier*, *Relation*, *Source*, *Language* und *Coverage* festgelegt. Die Suchmaschine extrahiert diese Daten und legt sie in ihrer Metadatenbank ab. Die o. g. Nachteile der automatischen Methoden kommen hierbei nicht zum tragen. Es ist eine Gewichtung möglich und durch das Ändern eines Dokumentes muss nicht zwangsläufig die Metadatenbank aktualisiert werden. Allerdings besteht die Möglichkeit der Manipulation durch den Autor.

### 2.3 Suchanfrage

Der Benutzer greift auf die Suchmaschine mit Hilfe eines Anfrageformulars zu. Die Stichwörter können logisch (AND, OR, NOT), als auch hinsichtlich ihres relativen Vorkommens (NEAR) kombiniert werden. Der Suchdienst generiert aufgrund der Suchbegriffe allein durch Rückgriff auf die Metadatenbank ein Antwortdokument, in

dem die Fundstellen als Hyperlinks dargestellt sind. Häufig wird auch eine Gewichtung sowie ein Textauszug angegeben.

### **2.4 Verteilte und mehrfach verteilte Suchmaschinen**

Aufgrund der Größe des World Wide Webs und dem damit verbundenen Umfang der zu indizierenden Dokumente finden bei Suchmaschinen vielfach Techniken aus dem Bereich der Verteilten Systeme Anwendung. Viele Suchdienste setzen mehrere Kopien ihrer Roboter ein, um das Internet schneller indizieren und dem Benutzer damit konsistentere Metadaten bieten zu können. Die Metadatenbank wird jedoch zentral auf dem physischen Plattenspeicher eines Supercomputers gespeichert. Hierbei werden jedoch häufig die Grenzen der Skalierbarkeit in bezug auf Größe und Performance erreicht.

Aus diesem Grund gehen Suchdienste dazu über, auch die Metadatenbank verteilt zu halten. Die Benutzerschnittstelle muss dann ihre Anfrage an alle Teile der Datenbank stellen. Damit erreicht man zwar eine nahezu beliebige Skalierbarkeit, was jedoch mit deutlichen Beschränkungen in der Ausdrucksstärke der Anfragesprache bezahlt wird.

### **2.5 Metasuchmaschinen**

Wie zu Beginn erwähnt, wurde in [Lawr98] festgestellt, dass ein Abfragen mehrerer Suchmaschinen den Suchraum erheblich vergrößern kann. Bereits die o. g. All-in-One-Formulare leiteten die Suchbegriffe an mehrere Suchmaschinen weiter.

Diese Idee wurde von den Metasuchmaschinen aufgegriffen. Der Benutzer definiert mittels eines Formulars die Suche, die dann an verschiedene herkömmliche Suchmaschinen weitergegeben wird. Die Metasuchmaschine selbst unterhält keine Metadatenbank. Durch die parallele Abfrage der Suchmaschinen wird ein erheblicher Performancegewinn erzielt. Sobald die Metasuchmaschine die einzelnen Ergebnisse erhalten hat, werden diese vor der Anzeige für den Benutzer weiter aufbereitet. Hierzu gehört u. a. die Integration der Ergebnisse, die Doubletten-Eliminierung, mit der die

Anzeige mehrerer Treffer auf das gleiche Dokument vermieden werden sollen, sowie die Sortierung nach Relevanz.

In [Sand98a] wurden sieben Kriterien für die Bewertung von Suchmaschinen formuliert:

- Die Metasuchmaschine muss wirklich parallel suchen.
- Die Ergebnisse müssen zusammengeführt und in einem einheitlichen Format dargestellt werden.
- Doppelte Fundstellen müssen erkannt und gekennzeichnet werden.
- Für logische Operationen müssen mindestens die Operatoren AND und OR zur Verfügung stehen.
- Wenn ein Suchdienst eine Kurzbeschreibung der Fundstelle liefert, denn muss diese übernommen werden.
- Die spezifischen Eigenschaften der unter der Metasuchmaschine liegenden Suchdienste dürfen für die Bedienung keine Rolle spielen, der Anwender muss nichts darüber wissen.
- Die Metasuchmaschine sollte in der Lage sein, so lange zu suchen, bis irgendeine der darunterliegenden Suchdienste noch Treffer liefert.

Die z. Zt. leistungsfähigsten Suchmaschinen erfüllen sechs der sieben Kriterien. Es sind der MetaCrawler, Highway61 und der auf deutsche Internet-Inhalte spezialisierte MetaGer<sup>6</sup>.

Neben den genannten Vorteilen haben auch Metasuchmaschinen verschiedene Nachteile. So ist trotz der parallelen Abfrage mit einer höheren Antwortzeit zu rechnen, da die Ergebnisse erst nach der Antwort der langsamsten Suchmaschine dargestellt werden können. Zusätzlich wird für die Übersetzung der Suchanfrage in das Format der jeweiligen Suchmaschine, sowie die abschließende Aufbereitung der Ergebnisse Zeit

---

<sup>6</sup> Die Adressen lauten: <http://www.metacrawler.com>, <http://www.highway61.com/> und <http://meta.rzrn.uni-hannover.de/>

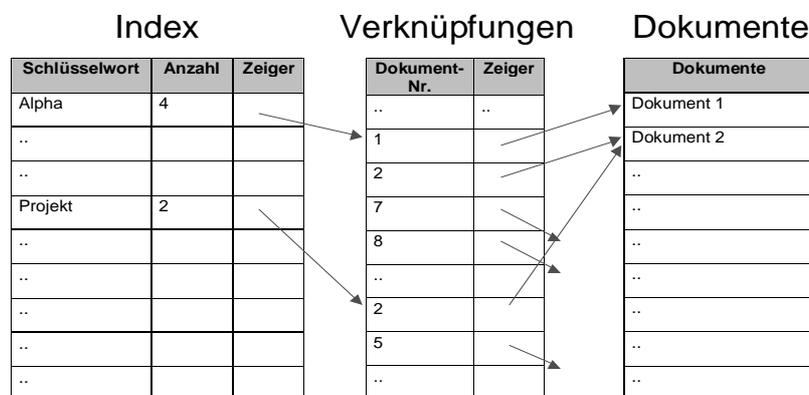
benötigt. Auch wird durch die parallelen Abfragen eine höhere Netzbelastung hervorgerufen. Die Nutzung erweiterter Suchoptionen der einzelnen Suchdienste ist nur eingeschränkt möglich.

### 3 Information Retrieval

Im folgenden wird auf einige grundlegenden Techniken des Information Retrieval eingegangen. Es wird die Datenstruktur der Inverted Files erläutert, sowie die Ausdrucks- bzw. Dokumentoperationen Stemming, Weighting / Ranking dargestellt. Die Anwendung von Thesaurus und Stoplists wird erläutert.

#### 3.1 Inverted Files

Eine effektive Datenstruktur für die Metadatenbank sind Inverted Files bzw. invertierte Dateien. Es handelt sich hierbei um einen lexikographischen Index, der eine schnelle Suche in großen Datenmengen erlaubt. Für diesen Index müssen zwischen 10% und 100% zusätzlicher Speicherplatz veranschlagt werden.



**Abbildung 2: Inverted File als sortierte Liste realisiert**

Die Datenstruktur Inverted File besteht aus einer sortierten Liste von Schlüsselwörtern (s. Abbildung 2). Diese enthält neben einer möglichen Gewichtung einen Zeiger auf das entsprechende Dokument oder auf eine Verknüpfungstabelle, die ihrerseits Zeiger auf die Dokumente enthält.

### 3.2 Stemming

Unter Stemming wird das Finden von morphologischen Varianten eines Begriffs verstanden. Dieses Verfahren kann sowohl auf die Suchbegriffe, als auch auf die zu indizierenden Wörter angewendet werden. Dadurch lässt sich eine Verringerung der Größe der Metadatenbank um bis zu 50% erreichen.

Neben dem manuellen Anlegen von Stemming-Tabellen können automatische Methoden eingesetzt werden. Beim *affix removal* werden Vor- und Nachsilben entfernt, um den Wortstamm zu erhalten, wohingegen *successor variety* auf struktureller Linguistik basiert, welche die Grenzen von Worten und Morphemen basierend auf der Verteilung von Silben in langen Texten zu erkennen versucht. Beim *n-gram* werden Suchbegriffe in 2-Buchstaben-Paare gebrochen und miteinander verglichen.

### 3.3 Ranking / Weighting

Bei der Suche in großen Informationsbeständen liefern Retrieval-Systeme häufig eine unüberschaubare Anzahl von Treffern. Diese Menge kann mit Hilfe von logischen Operationen eingeschränkt werden, wobei die Anwendung dieser Methoden den gelegentlichen Benutzer von Suchmaschinen im Internet überfordern kann, da er sich erst in die Anfragesyntax der jeweiligen Suchmaschine einarbeiten muss.

Mit Hilfe der Rangfolge- und Gewichtungsalgorithmen versuchen Suchmaschinen, dem Benutzer einen Hinweis auf die Relevanz der Fundstelle zu geben und diese bei der Anzeige entsprechend zu sortieren. Hierdurch können auch lange Trefferlisten sinnvoll genutzt werden und es kann weitestgehend auf den Einsatz von logischen Operationen verzichtet werden. Damit ergibt sich neben einer einfacheren Benutzerschnittstelle eine gewisse Fehlertoleranz der Suchmaschine.

Eine möglicher Gewichtungsalgorithmus könnte wie folgt vorgehen:

- Der Informationsbestand besteht aus  $i$  unterschiedlichen Ausdrücken.
- Ein Dokument  $j$  wird durch einen Vektor  $T_j=(t_1, t_2, \dots, t_n)$  repräsentiert mit:

- $t_i = 1$ , wenn der Ausdruck  $i$  im Dokument  $j$  vorkommt
- $t_i = 0$ , sonst.
- Die Suchanfragen werden durch einen Vektor  $q$  in der gleichen Weise repräsentiert.

Das Kreuzprodukt von  $T_j$  und  $q$  wird hierbei als Vergleichswert für die Relevanz des Dokumentes verwendet.

Eine zusätzliche Gewichtung könnte der Vektor  $T_j$  durch folgendes Bildungsgesetz erhalten:

- $t_i = k$ , wenn der Ausdruck  $i$  im Dokument  $j$  mit der Häufigkeit  $k$  vorkommt.

### 3.4 Thesaurus

Ein Thesaurus „ist ein kontrolliertes, dynamisches Vokabular von bedeutungsmäßig und generisch verbundenen Termini, das einen spezifischen Fachbereich umfassend abdeckt“ (DIN 1463). Einerseits kann er den Benutzer bei der Formulierung einer präziseren Suche unterstützen, da dieser einen treffenderen Ausdruck für seinen Suchbegriff nachschlagen kann. Andererseits kann die Suchmaschine mit Hilfe des Thesaurus sowohl ein allgemeineres, als auch ein präziseres Suchergebnis erstellen.

Als Beispiel sei der Eintrag im Thesaurus für den Begriff „computer-aided instruction“ gegeben. Ein alternativer Ausdruck ist „teaching machines“. Ein allgemeinerer Begriff ist „educational computing“, wobei „computer application“ der Oberbegriff hierzu ist. „Education“ und „teaching“ stehen in Bezug zu „computer-aided instruction“.

### 3.5 Stoplists

Um den Speicherbedarf für die Metadatenbank zu vermindern, werden Wörter, die extrem häufig vorkommen oder für den Index nicht benötigt werden, in Stoplists bzw. Stoppwortlisten eingetragen und nicht indiziert. Dieses können Artikel, Konjunktionen o. ä. sein.

Ein Liste von 425 Wörtern für englische Texte [Fran82] beginnt wie folgt:

a	about	above	across	after
again	against	all	almost	alone
along	already	also	although	always
among	an	and	another	any
anybody	anyone	anything	anywhere	are
area	areas	around	as	ask
asked	asking	asks	at	away
b	..	..	..	..

**Tabelle 1: Beispiel einer Stoplist**

## 4 Fazit

Die Ausführungen haben gezeigt, dass inzwischen eine Vielzahl von Anstrengungen unternommen worden sind, um das im Internet verfügbare Wissenspotential zu erschließen. Gleichzeitig existieren aber noch viele Probleme, deren Lösung noch nicht erfolgt ist.

### 4.1 Probleme existierender Suchhilfen

Das schnelle Veralten der Metadatenbank sei hier als Erstes genannt, da hiervon alle Suchmaschinen und damit auch die föderierten Suchdienste, wie die Metasuchmaschinen betroffen sind. Durch die Dynamik und das schnelle Wachstum des World Wide Webs wird sich dieses Problem noch verstärken.

Hiermit verbunden ist auch die Problematik der Duplikate, die von Dokumenten im Internet, meist ohne Wissen des Autors angelegt werden. Eine Unterscheidung zwischen Original und Kopie ist für Suchmaschinen nicht möglich. Ebenso können bei einer Aktualisierung des Originals die Kopien veralten. In diesem Zusammenhang sind auch die Phantomkopien zu erwähnen. Hierbei handelt es sich um mehrere logische Adressen für ein Dokument.

Wie bei herkömmlichen Publikationen ist eine Zuordnung der verbreiteten Dokumente zu Autoren grundsätzlich wünschenswert und ein wichtiges Metadatum für Suchmaschinen. Häufig sind Dokumente nur für eine bestimmte Leserschaft vorgesehen. Suchmaschinen können durch eine Hinweisdatei von der Indizierung bestimmter Dokumente abgehalten werden.

Obwohl Englisch die im Internet vorherrschende Sprache ist, müssen von Suchmaschinen auch andere Sprachen berücksichtigt werden. Spielt dieses beim Aufbau der Metadaten zunächst eine untergeordnete Rolle, wird das Problem aber im Zusammenhang mit Stoplists oder einem Thesaurus offenkundig. Viele Suchmaschinen haben inzwischen europäische oder deutsche Ableger gegründet, die sich auf die Indizierung in einem bestimmten Sprachraum spezialisiert haben<sup>7</sup>.

Auch wenn heute multimediale Elemente noch weitestgehend von Suchmaschinen ausgeschlossen werden, so wird die Berücksichtigung dieser Informationen im Hinblick auf die zunehmende Verbreitung dieser Inhalte im World Wide Web immer wichtiger.

### **4.2 Nutzen für Electronic Commerce**

Derzeit kann von Suchmaschinen kein Nutzen für den elektronischen Handel erwartet werden. Wie die vorangegangenen Darstellungen gezeigt haben, können Suchmaschinen lediglich Dokumente indizieren, die sie durch einen Hyperlink eines anderen Dokuments erreicht haben. In Systemen des Electronic Commerce werden jedoch fast ausschließlich dynamisch generierte Informationen erzeugt, die nicht über einen Verweis zu erreichen sind, sondern von einem CGI-Programm<sup>8</sup> erzeugt werden. Damit bleiben diese Dokumente den Suchmaschinen verschlossen. Aufgrund der dezentralen und autarken Organisation des World Wide Webs sind an dieser Stelle auch kaum Eingriffe denkbar bzw. möglich.

---

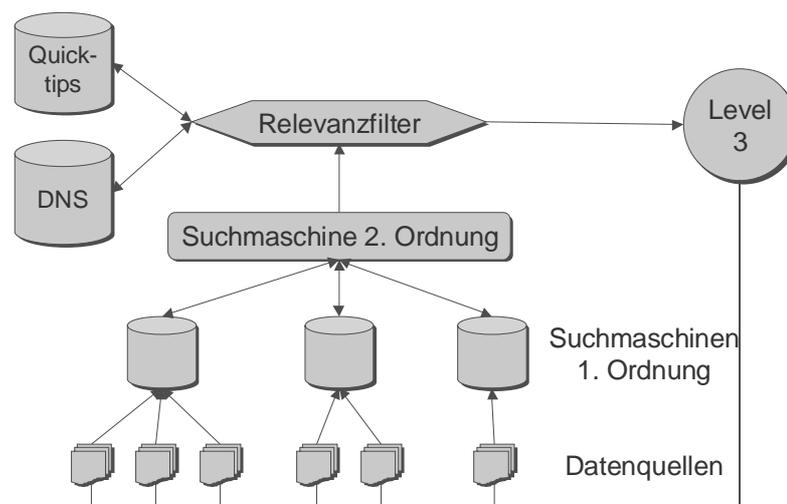
<sup>7</sup> Als Beispiel sei hier der deutsche Ableger von Lycos erwähnt: <http://www.lycos.de>

<sup>8</sup> CGI (Common Gateway Interface) bietet eine programmierbare Schnittstelle des Servers, mit deren Hilfe auf Benutzereingaben reagiert werden kann.

Ein Lösungsansatz an dieser Stelle wäre, eine standardisierte Schnittstelle zwischen E-Commerce-Anbieter und speziell hierfür vorgesehenen Suchmaschinen zu schaffen. Über diese Schnittstelle könnten Informationen über Waren und Preise ausgetauscht werden. Ein Benutzer einer solchen Suchmaschine könnte dann einen Anbieter suchen lassen, der seinen Produkt- und Preisvorstellungen entspricht. Dieses Vorgehen weist Ähnlichkeiten mit der Technologie der autonomen Agenten auf.

### 4.3 Ausblick

Unabhängig von der Verwendbarkeit für Electronic Commerce wird die Entwicklung der Suchmaschinen weiter verfolgt. Einen weiteren Nutzensgewinn bringt die in [Sand98b] diskutierte Suchmaschine der nächsten Generation (s. Abbildung 3). Diese baut auf einer Metasuchmaschine auf. Neben einer Relevanzfilterung sieht sie auch noch den Einsatz von QuickTips und der DNS-Datenbank vor.



**Abbildung 3: Metasuchmaschine der nächsten Generation**

Bei den QuickTips handelt es sich um einen manuell aufgebauten Katalog ausgewählter Adressen. Durch eine Anfrage an Domain-Name-Server (DNS) wird versucht, zu einem Suchbegriff einen Domänennamen zu finden, da die Wahrscheinlichkeit, auf dem Server einer solchen Domäne relevante Informationen zu finden, außerordentlich hoch sein kann.

## Literaturverzeichnis

- [Dubl97] Dublin Core Element Set ([http://purl.org/metadata/dublin\\_core\\_elements](http://purl.org/metadata/dublin_core_elements), 24.11.1998).
- [Frak92] *Frakes, W. B.; Baeza-Yates, R.*: Information Retrieval – Data Structures & Algorithms, Prentice-Hall, London 1992.
- [Fran82] *Frances, W.; Kucera, H.*: Frequency Analysis of English Usage, Houghton Mifflin, New York, 1982.
- [Jeus97] *Jeusfeld, M. A.; Jarke, M.*: Suchhilfen für das World Wide Web: Funktionsweisen und Metastrukturen. In: *Wirtschaftsinformatik*, 39 (1997) 5, S. 491-499
- [Keit93] *Keitz, S. von; Keitz, W. von; Gerlach, H.*: Modernes Online Retrieval, VCH, Weinheim, 1993.
- [Lawr98] *Lawrence, S.; Giles, C. L.*: Searching the World Wide Web. In: *Science*: 1998 April 3; 280 (5360): 98.
- [Mase97] *Masermann, U.*: Suchdienste im World-Wide Web – Anforderungen und Perspektiven. In: *Jeusfeld, M. (Hrsg.)*: Informationssysteme für das Internet – Anforderungen, Konzepte, Methoden. Proceedings EMISA-Fachgruppentreffen 1996 [EMISA-FORUM (1997) 1], Gesellschaft für Informatik, 1997.
- [Sand98a] *Sander-Beuermann, W.; Schonburg, M.*: Internet Information Retrieval – The Further Development of Meta-Searching Technology. Proceedings of the Internet Summit, Internet Society, 1998, Genf (<http://www.uni-hannover.de/inet98/paper.html>, 24.11.1998).
- [Sand98b] *Sander-Beuermann, W.*: Schatzsucher – Die Internet-Suchmaschine der Zukunft. In: *c't Magazin für Computer Technik*, (1998) 13, S. 178-184.
- [Teut97] *Teuteberg, F.*: Effektives Suchen im World Wide Web: Suchmaschinen und Suchmethoden. In: *Wirtschaftsinformatik* 39 (1997) 4, S. 373-383.
- [Voss98] *Vossen, G.; Masermann, U.*: Suchmaschinen und Anfragen im World Wide Web. In: *Informatik Spektrum* 21 (1998), S. 9-15.