

<b>1. EINFÜHRUNG INS INTERNET</b>	<b>2</b>
1.1. WAS IST DAS INTERNET?	2
1.2. GESCHICHTE DES INTERNET	3
1.3. STRUKTUR DES INTERNET	4
1.3.1. DAS INTERNET IN ZAHLEN	4
1.3.2. INTERNETDIENSTE	4
1.4. WIE FUNKTIONIERT DAS INTERNET?	6
1.4.1. GRUNDLAGEN	6
1.4.2. WAS SIND PROTOKOLLE?	6
1.4.3. HTTP	7
1.4.4. ADRESSEN IM INTERNET	7
1.4.5. AUFBAU VON WWW-ADRESSEN (URL)	8
<b>2. SUCHMASCHINEN, KATALOGE UND PORTALE</b>	<b>10</b>
2.1. SUCHMASCHINEN	10
2.2. KATALOGE	12
2.3. METASUCHMASCHINEN	13
2.3.1. KATEGORIEN VON SUCHMASCHINEN	14
2.4. PORTALE	14
<b>3. INFORMATIONEN SUCHEN UND FINDEN</b>	<b>16</b>
3.1. SCHNELLES AUFFINDEN AKTUELLER INFORMATIONEN	16
3.2. SCHNELLES AUFFINDEN UMFANGREICHER INFORMATIONEN	FEHLER! TEXTMARKE NICHT DEFINIERT.
3.3. DIE LANGFRISTRECHERCHE	FEHLER! TEXTMARKE NICHT DEFINIERT.
3.4. DER RICHTIGE EINSATZ VON SUCHMASCHINEN	16
3.4.1. DIE STICHSUCHTE	16
3.4.2. DIE „RICHTIGE“ FRAGETECHNIK	FEHLER! TEXTMARKE NICHT DEFINIERT.
3.4.3. SUCHTECHNIKEN ANWENDEN	19
3.5. UNTERSCHIEDLICHE SUCHMASCHINEN UND IHRE BEDIENUNG	24

## 1. Einführung ins Internet

### 1.1. Was ist das Internet?

Der Begriff setzt sich zusammen aus:



⇒ „Zwischennetz“ (Netz der Netze)

Das Internet verbindet weltweit einzelne, voneinander unabhängige Rechnernetze. Dadurch ist der weltweite Datenaustausch und die weltweite Kommunikation möglich

### 1.2. Geschichte des Internet

- 1969 **ARPANET**(Advanced Research Project Agency)  
Netzwerk des amerikanischen Verteidigungsministeriums
- 1970 **E-Mail:**  
Erstes E-Mail-Programm für gemischte Netzwerke
- 1982 **TCP/IP**  
Einführung des Netzwerkprotokolls TCP/IP  
TCP/IP wird im weiteren Verlauf weltweiter Standard bei der Vernetzung von Unix-Rechnern.
- 1983 **ARPANET/ MILNET**  
Aufteilung des ARPANET in den militärischen Teil (MILNET) und den nicht-militärischen (ARPANET).
- 1984 **DNS**  
Domain-Name-Server (DNS) werden eingeführt.
- 1989 Erste Schnittstelle zwischen einem kommerziellen Mailbetreiber und dem Internet
- 1989 **INTERNET**  
ARPANET beendet seine Arbeit. Fortan wird das gesamte Netz INTERNET genannt.
- 1990 Erster kommerzieller Internet-Provider in den USA
- 1990 **WWW**  
Der Dienst **World Wide Web** wird der Öffentlichkeit vorgestellt.
- 1993 Mit **Mosaic**, dem ersten grafikfähigen Web-Browser, beginnt der Siegeszug des WWW = World Wide Web um die Welt.
- Benutzerfreundliche Anwendungsprogramme
  - Multimediale Unterstützung (Grafiken, Videos, Musik, Sprache, Text)
  - Durch das WWW wird das Medium Internet populär
  - Fast jedes Jahr verdoppelte sich die Zahl der Server
- 1991 **IV - DE - NIC**  
Netzwerk Informations Centrum (NIC) zur Verwaltung der Internet-Namen mit der Endung \*.de wird gegründet. (DE ist das internationale Kürzel für Deutschland).
- 1994 ARPANET/ INTERNET feiert seinen 25. Geburtstag.  
Erste Online-Einkaufsmöglichkeiten
- 1999 Das INTERNET feiert seinen 30. Geburtstag.

### 1.3. Struktur des Internet

- weltweite Verbreitung
- dezentrale Struktur, keine „Hierarchie“ und keine „Zentralverwaltung“, die das ganze Netz kennt
- Transmission Control Protocol/ Internet Protocol (**TCP/IP**):
- Mit dem TCP/IP Protokoll liegt ein einheitlicher Datenübertragungsstandard vor, der **hardwareunabhängig** ist. Damit ist der Zusammenschluss von Computern verschiedener Bauarten und Hersteller möglich.
- Zusammenschluss von Providern, um den **providerübergreifenden Datenverkehr** zu beschleunigen
- Teilnetze einzelner Provider liegen nebeneinander
- mangels internationalen Abkommen und Vollstreckbarkeit nur eingeschränkt kontrollierbar
- unzensuriert, freier Zugang, aber: Beschränkung wird zunehmend diskutiert (vgl. [www.eff.org](http://www.eff.org) und [www.catalaw.com/doom](http://www.catalaw.com/doom))
- Online-Dienste bieten Internet-Zugang über eigene Netztopologie an, z.B. T-Online plus Internet

#### 1.3.1. Das Internet in Zahlen

- Ca. 44 Mio. Rechner
- Ca. 163,5 Mio. Anwender weltweit
- Ca. 36 Mio. Anwender in Europa
- Ca. 7,5 Mio. Anwender in Deutschland
- Ca. 1,2 Mio. Anwender in der Schweiz
- Ca. 0,5 Mio. Anwender in Österreich
- Ca. 2 Mio. neue Anwender pro Monat
- 11 % der deutschen Internet-Anwender kaufen bereits über das Internet ein.
- Im Jahre 2002 werden für 1,4 Mrd. DM Flugreisen und für 1,2 Mio. DM Bücher umgesetzt.

#### 1.3.2. Internetdienste

Als **Dienst** wird im Internet **eine Sammlung von Funktionen oder Fähigkeiten, die zur Nutzung bereitgestellt werden** bezeichnet.

**Voraussetzung** für die Nutzung eines Dienstes:

- Dienstanbieter (z.B. T-Online oder AOL)
- Hardware
- Software

- **World Wide Web (WWW)**

Multimediale Nutzungsmöglichkeit des Internet zur Informationsbeschaffung, für Werbung, On-line-Shopping etc.

Anwenderfreundlich durch grafische Benutzeroberfläche der Anwendungsprogramme

- **E-Mail**

Elektronische Post

Empfangen und Versenden von elektronischen Nachrichten bzw. Informationen

- **Newsgroups (Diskussionsforen)**

Öffentliches „schwarzes Brett“

Nachrichten können von jedermann gelesen oder „angehängt“ werden

- **FTP (File Transfer Protocol)**

Übertragung und Kopieren von Dateien zwischen zwei Rechnern, zum Download von Software oder Updates

- **IRC (Internet Relay Chat)**

Direkte Kommunikation verschiedener Internet-Nutzer in Echtzeit

- **Internettelefonieren**

- Telefongespräche direkt im Internet. Noch sehr schlechte Sprachqualität.
- Mittlerweile durch billig gewordene „normale“ Telefongespräche uninteressant.

### 1.4. Wie funktioniert das Internet?

#### 1.4.1. Grundlagen

Kommunikation im Internet erfolgt nach dem *Client-Server-Prinzip*:

- Der Client (=PC am Arbeitsplatz) fordert eine Dienstleistung an.
- Der Server (=Internet-Rechner des Anbieters) bietet eine Dienstleistung an.

Für die Kommunikation zwischen verschiedenen Rechnern mit unterschiedlichen Betriebssystemen sind Regeln erforderlich (*Protokolle*).

#### 1.4.2. Was sind Protokolle?

*Regeln zur Verständigung von Rechnern untereinander in einem Netzwerk*

Um über ein Netzwerk zu kommunizieren, müssen Rechner die gleiche „Sprache“ sprechen, d.h. die gleichen Protokolle verwenden. Es existieren verschiedene Protokolle zur Erfüllung unterschiedlicher Aufgaben (Transport, Überwachung, Wegoptimierung, Darstellung etc.). Das übergeordnete Protokoll, auf dem andere Protokolle im Internet aufbauen, ist das *TCP/IP Protokoll*.

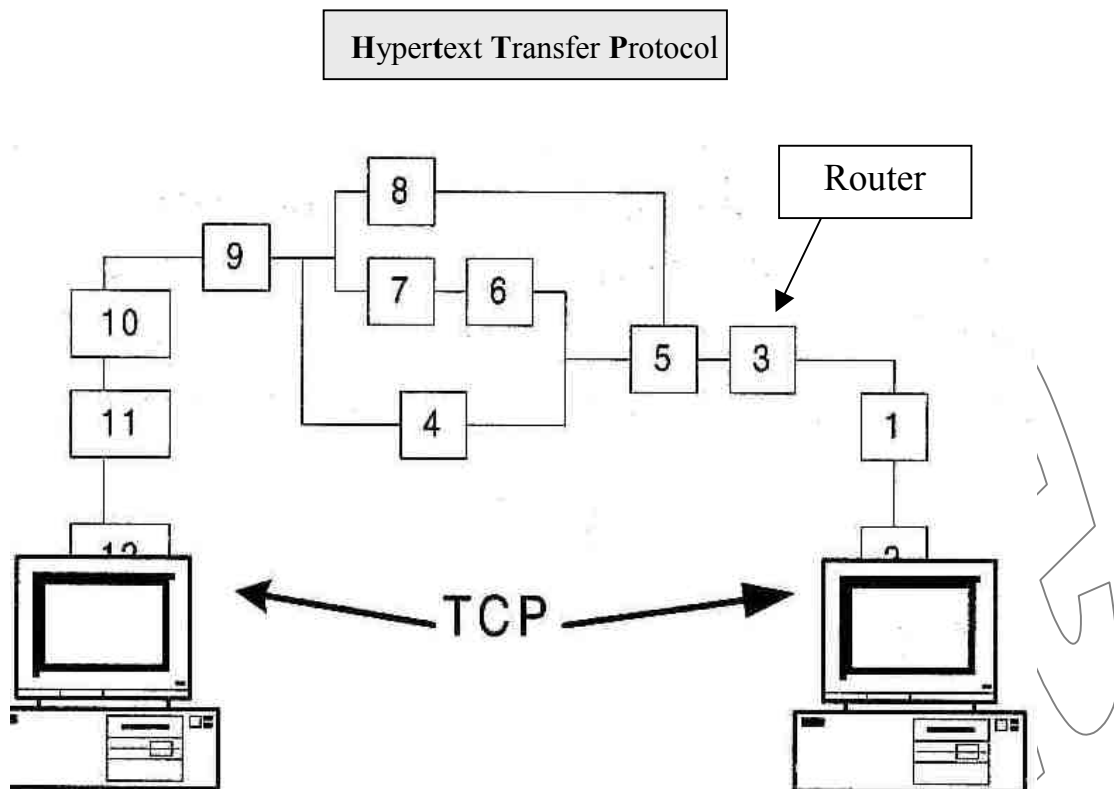
**TCP = Transmission Control Protocol**  
**IP = Internet Protocol**

Die beiden Protokolle arbeiten eng zusammen und bauen aufeinander auf, abgekürzt: **TCP/IP**.

**TCP/IP legt fest, wie Daten in einem Netzwerk adressiert und versendet werden.**

- Die Daten werden in gleich große Pakete zerlegt und nacheinander versendet.
- Jedes Paket erhält die Absender- und die Empfängeradresse.
- Das Paket wird über Router im Netz weitergeleitet.
- Datenpakete dürfen maximal ca. 1500 Zeichen groß sein, damit eine gleichmäßige Netzbelastung gewährleistet werden kann.
- Datenpakete können über verschiedene Wege innerhalb des Netzes vom Sender zum Empfänger gelangen.
- Beim Empfänger wird die Nachricht aus den Teilnachrichten wieder zusammengesetzt.
- Anhand einer Prüfsumme kann erkannt werden, ob die übertragenen Daten fehlerhaft sind. Beim Auftreten eines Fehlers wird eine erneute Sendung des betreffenden Pakets eingeleitet.

### 1.4.3. HTTP



Dieses Protokoll regelt die Verbindung und die Kommunikation zwischen WWW-Client und WWW-Server. Das **Unterprotokoll HTTP** setzt auf TCP/IP auf. Es wurde speziell für das Arbeiten mit Hypertext-Dokumenten im Internet entwickelt. Hypertext-Dokumente bilden das WWW.

### 1.4.4. Adressen im Internet

Jeder Rechner im Internet besitzt eine eindeutige Adresse, die sogenannte IP-Adresse (IP = Internet Protocol). Diese IP-Adresse kann numerisch oder symbolisch dargestellt werden.

- **Numerische Adresse (IP-Adresse)**

Numerische Adressen (32 Bit)

z.B. 178.023.128.67 oder 164.154.171.187

IP-Adresse besteht aus zwei Teilen:

Netzwerkadresse (Net-ID)

Rechneradresse (Host-ID)

Allgemeine Darstellung:

Netzwerkadresse.Rechneradresse

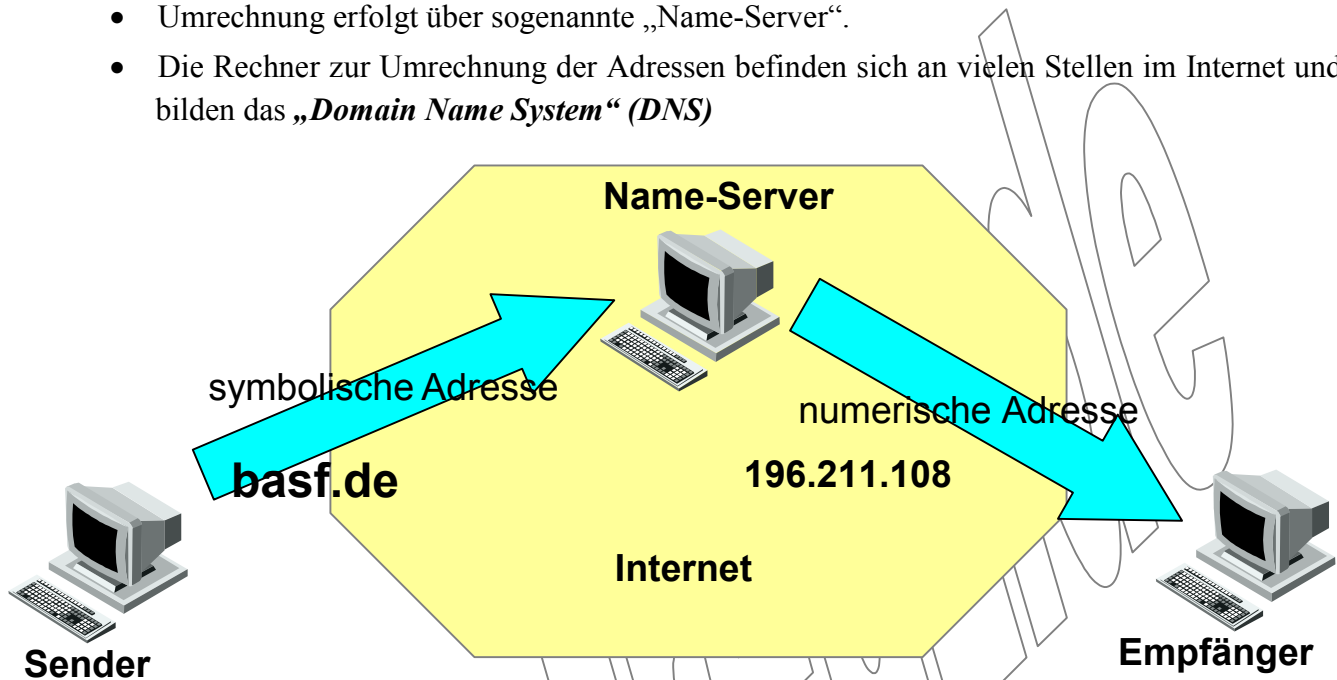
- **Symbolische Adresse (IP-Adresse)**

Weil numerische Adressen schwer zu behalten sind, werden sie durch symbolische Adressen ersetzt.

Beispiel: 194.97.68.1 = basf.de

Das Netz wird zur eindeutigen Adressierung in Bereiche sogenannte Domains aufgeteilt. Diese Bereiche sind hierarchisch organisiert.

- Ein Rechner im Internet kann über die numerische oder die symbolische Adresse angesprochen werden.
- Die symbolische Adresse wird automatisch in eine numerische umgerechnet.
- Umrechnung erfolgt über sogenannte „Name-Server“.
- Die Rechner zur Umrechnung der Adressen befinden sich an vielen Stellen im Internet und bilden das „*Domain Name System*“ (DNS)



### 1.4.5. Aufbau von WWW-Adressen (URL)

URL = *Uniform Resource Locator*

Die URL's sind ein *symbolisches Adressierungsschema*. Die URL lokalisiert:

- Internetdienst (z.B. www oder ftp)
- Server
- Ort der gewünschten Information
- Darstellung:
- Protokoll://Servername.Domain.TopLevelDomain/Verzeichnis/Dateiname
- *Aufbau der symbolischen Adressen*

Servername.[LocalDomain].SecondLevelDomain.TopLevelDomain

www.wi.uni-muenster.de



*TopLevelDomains im Überblick:*

- .com = Kommerzielle Anbieter
- .edu = Bildungseinrichtungen
- .mil = Militär
- .gov = nicht-militärische Regierungsangelegenheiten
- .net = Netzwerke
- org = andere Organisationen

## 2. Suchmaschinen, Kataloge und Portale

„Wer sucht, der findet auch“. Es gibt so gut wie nichts, was nicht im Internet steckt. Oder auch versteckt wäre. Die Suche in der Menge an Dateien, Bildern und Dokumente ist vergleichbar mit der Suche einer Stecknadel im Heuhaufen, denn aufgrund der erfreulich freien und offenen Struktur des Internets gibt es keine Zentralinstanz, die Angebote sammelt und systematisch verwaltet. So ist die Suche nach speziellen Informationen oft mühsam, aufwendig und für den Benutzer auch kostenintensiv. Schneller zum Ziel kommt man mit dem Nutzen eines geeigneten Werkzeuges: Die *Suchmaschinen* und die *Web-Kataloge*

### 2.1. Suchmaschinen

*Suchmaschinen* sind *leistungsfähige Datenbanken*, in denen riesige abfragbare Datenbanken erstellt werden. Eine Volltextsuchmaschine besteht im wesentlichen aus drei Teilen:

- aus Datensammlern,
- einem Indizierungsprogramm
- sowie einem Abfrageprogramm.

Die *Datensammler*, *Robots* genannt, besuchen rund um die Uhr Websites, rufen also die betreffenden HTML-Seiten ab. Dabei setzen die Suchmaschinen viele Robots gleichzeitig ein, manche schicken Hunderte davon los. Nur so können die Recherchehelfer überhaupt mehrere Millionen Webseiten täglich indizieren.

**Spider, Crawler, Robots** - alle drei Begriffe haben gemeinsam, dass sie Programme bezeichnen, die selbständig das World Wide Web nach Informationen durchsuchen.

Auf einer WWW-Seite sind meist Verknüpfungen (=Links) angegeben, die zu weiteren Seiten führen, auf denen wiederum Verknüpfungen vorhanden sind. An diesem Vorgang erkennen Sie auch, wie eine solche Suchmaschine grundsätzlich arbeitet. Sie erhält die Grundinformation über eine Reihe von Internet-Seiten und macht sich dann an die Arbeit, d.h., jeder Link auf einer Seite wird festgehalten und dann ebenfalls durchsucht. Von dort aus geht es wieder weiter - vollautomatisch! Ein solcher **Webcrawler** braucht nur einmal angestoßen zu werden, dann arbeitet er sich wie ein Roboter **automatisch** durch das gesamte Internet hindurch, wodurch sich auch der Name **Robot** erklärt.

Allerdings sind die Robots der Suchmaschinen so ausgelastet, dass es eine Weile dauern kann, bis ein Suchhelfer eine neue Seite in den Index aufnimmt oder die Veränderung einer Website wahrnimmt. Bei AitaVista, Fireball und alltheweb etwa dauert es nur ein bis zwei Tage, bis eine neu angemeldete URL im Index erscheint. Ander Suchmaschinen können hierfür bis zu sechs Monate benötigen. Wird der Robots auf eine Site angesetzt, so sammelt er oft nicht alle zugehörigen Seiten, sondern nimmt nur eine Stichprobe.

Wie viele Seiten ein Robot jeweils indiziert, variiert sehr stark von Suchmaschine zu Suchmaschine. Einige Elemente einer Website können Robots daran hindern, darin verlinkte Dokumente

zu besuchen. So verfolgen beispielsweise Excite und Lycos keine Links, die sich hinter Image-maps verbergen; Inktomi und Go unterstützen keine Frames. Webmaster können Robots aber auch explizit verbieten, ihre Sites zu indizieren.

Die von den Robots ermittelten Daten verarbeitet der **Indizierer** weiter. Abhängig von den in den Seiten enthaltenen Wörtern baut er einen Volltextindex auf. Wie die Indizierer Seiten priorisieren (**ranking**), also in Bezug zu bestimmten Suchbegriffen im Index einstuft, hängt von verschiedenen Faktoren ab. So spielt es eine wichtige Rolle, wie häufig und an welcher Stelle der Suchbegriff im Dokument vorkommt: In der URL, im Titel, in Überschriften, in normalen Absätzen oder in so genannten Meta-Tags.

Allerdings gibt es **Unterschiede in der Arbeitsweise** der Robots, und zwar entscheidende! Manche dieser Roboter sammeln nur die einzelnen Links und Überschriften der Webseiten und erfassen sie in einer Datenbank. Andere Roboter erfassen den gesamten Text (Volltextrecherche) und arbeiten ihn in eine eigene Datenbank ein. Um Missverständnissen vorzubeugen: Nur diese Datenbanken werden im Moment ihrer Abfrage durchsucht, wenn Sie eine Suchmaschine aktivieren, nicht das gesamte Web! Wenn Sie also eine Suchabfrage starten, wird die Datenbank der Suchmaschine abgefragt - und Sie erhalten das Ergebnis.

Ein Kriterium, das immer mehr Suchmaschinen mit berücksichtigen, sind **externe Links** - je mehr andere Webangebote auf eine Site verweisen, desto besser wird diese bewertet. Damit berücksichtigen die Suchmaschinen gewissermaßen eine Vorbewertung durch die Internet-Gemeinde: Ist eine Site populär, so verweisen viele andere Seiten darauf. Vorreiter Google setzt sehr stark auf externe Links: Neben der Anzahl der auf eine Site zeigenden Verweise bewertet Google auch die Relevanz - Links von als wichtig eingestuften Sites wiegen mehr - und den Inhalt der Verweise. Neben Google benutzen auch AltaVista, Excite, Go und Inktomi externe Links zur Bewertung.

Aber auch das **Benutzerverhalten** lässt sich für das Ranking auswerten. Wer sich die Verweise in der Ergebnisliste von HotBot genauer ansieht, wird feststellen, dass diese nicht mit den angezeigten URL's übereinstimmen. HotBot benutzt vielmehr einen Redirect: Der Benutzer teilt zunächst, indem er auf den Verweis klickt, der Suchmaschine mit, auf welche Site des Ergebnisses er wechselt und wird dann erst von HotBot zum Ziel geführt. HotBot kann durch den Trick sogar die ungefähre **Verweildauer** der Surfer auf den Ergebnisseiten abschätzen. Verweise, die von Besuchern des Recherchehelfers öfter angeklickt werden und auf denen sie länger verweilen - offenbar populäre Treffer - bewertet HotBot höher.

Die **Ergebnisse** werden bei den einzelnen Suchmaschinen sehr unterschiedlich ausfallen. Jede Suchmaschine startet ihren „Spaziergang“ durch das Web von anderen Seiten aus - bei AltaVista sollen es zum Beispiel 3.000 Seiten sein, bei denen der Crawler zu suchen beginnt. Zum anderen ist die von einer Seite erfasste Datenmenge von unterschiedlicher Qualität. Manche Crawler erfassen nur den Namen der Webseite, manche nur den Anbieter der Webseite. Manche erfassen zusätzlich noch den ersten Absatz, weil man dort meist eine Zusammenfassung vermutet. Andere

Suchmaschinen erfassen die komplette Seite einschließlich aller Bestandteile, wie Bild-, Audio- und Videodateien.

Selbst wenn alle Daten identisch erfasst würden, bliebe noch **die Verarbeitung der Daten in der hauseigenen Datenbank**. Nach welchen Kriterien werden die Daten angeordnet und damit wieder auffindbar gemacht? Diese Frage beantworten alle Suchmaschinen unterschiedlich, und deshalb sind die Suchergebnisse auch sehr unterschiedlich.

Inzwischen gibt es sogar Suchmaschinen, die nur noch die Datenbanken einer Reihe von Suchmaschinen durchsuchen, sogenannte **Meta-Suchmaschinen**. Das ist zwar nicht uninteressant, wenn man die verfügbare Datenmenge berücksichtigt, aber es kostet eben doch zusätzlich mehr oder weniger viel Zeit.

## 2.2. Kataloge

**Kataloge (Verzeichnisse)** sind eine Besonderheit, d.h., diese Web-Verzeichnisse schicken keine Roboter auf die Suche, sondern dort **recherchiert** eine mehrköpfige Redaktion. Diese besucht Webseiten, rezensiert sie und sortiert sie in einen Schlagwortkatalog ein. Die Rubriken in diesen Katalogen sind hierarchisch gegliedert, die Links werden sortiert, mit einer kurzen Beschreibung versehen und inhaltlich passenden Themengebieten zugeordnet. Somit entstehen Inhaltsverzeichnisse ähnlich denen in einem Buch. Die Qualität eines Kataloges ist also wesentlich von der redaktionellen Kompetenz der Betreiber abhängig.

Verzeichnisse bieten keine Möglichkeit, wie Suchmaschinen den Inhalt der katalogisierten Sites zu durchsuchen. Yahoo, Web.de und Co. bieten zwar allesamt auch eine Volltextrecherche-funktion, aber diese durchsucht nur die im Katalog enthaltenen Beschreibungen der Sites oder liefert die Treffer einer Partner-Volltextsuchmaschine. Auch bei der Bewertung der Sites tun sich Kataloge schwer. **Typischerweise sind alle Verweise innerhalb einer Verzeichniskategorie gleichwertig und alphabetisch sortiert**; der Besucher kann höchstens noch erwarten, dass die Redakteure pro Rubrik ein paar Favoriten gekürt haben.

In der täglichen Praxis werden Sie **Unterschiede** zwischen Katalogen und Suchmaschinen erkennen können. Die Qualität der angebotenen Adressen ist bei Web-Katalogen erheblich höher, die Zahl der ermittelten Adressen ist wiederum bei den reinen Suchmaschinen höher.

**Die Unterschiede zwischen Web-Katalogen und Suchmaschinen verschwimmen aber immer mehr.**

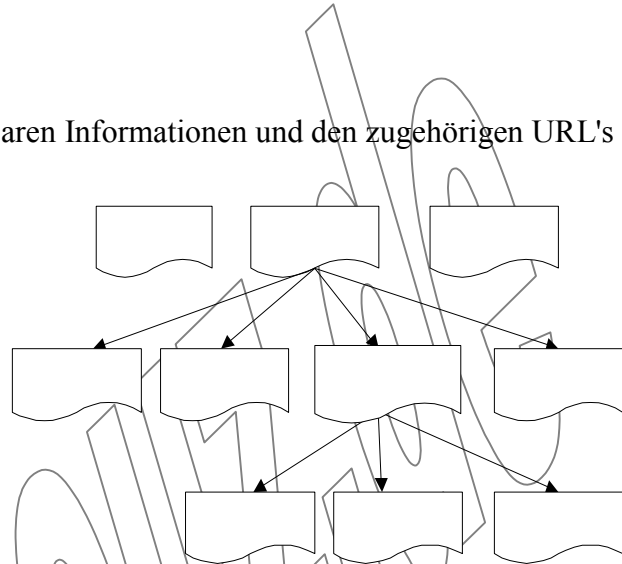
Die Web-Kataloge bieten zusätzliche Suchmaschinen und Informationen an, und die Suchmaschinen bieten die Suche nach Kategorien an, die redaktionell betreut werden. Manche Suchmaschinen machen dies in eigener Regie, andere vergeben dies an Zeitungen und Zeitschriften.

Unter den Betreibern sowohl von Volltextsuchern als auch von Web-Katalogen hat es sich herumgesprochen, dass sich die beiden unterschiedlichen Formen von Recherchediensten gut er-

gänzen. So treten immer mehr Suchdienste kombiniert auf. In Deutschland kooperiert beispielsweise Web.de mit AltaVista.de. Findet Web.de einen Suchbegriff nicht im eigenen Katalog, so leitet der Dienst die Anfrage an AltaVista weiter und präsentiert deren Ergebnisse. Umgekehrt ermöglicht die Volltextsuchmaschine auf ihrer Homepage den Zugriff auf die Rubriken von Web.de.

### Zusammenfassung:

- **Suchmaschinen (Search Engines)**
  - Datenbanken mit Stichwörtern zu verfügbaren Informationen und den zugehörigen URL's
- **Kataloge**
  - Thematisch sortierte Listen
  - Hierarchisch aufgebaut



Neben den allgemeinen Suchmaschinen und Katalogen gibt es noch Branchenbücher und spezielle Suchmaschinen.

**Branchenbücher** sind Suchmaschinen und Kataloge, die z.B. auf Telefonnummern, Adressen u. ä. spezialisiert sind.

**Spezielle Suchmaschinen** betreiben Datenbanken für bestimmte Bereiche wie Software und Newsgroups.

### 2.3. Metasuchmaschinen

Metasuchmaschinen bieten die Möglichkeit, gleich in mehreren Katalogen und Volltextsuchmaschinen zu recherchieren. Dabei leiten Metasuchmaschinen Anfragen an verschiedene Recherchehelfer weiter, sammeln deren Ergebnisse und bereiten sie in einer einheitlichen Weise auf - da der Benutzer mit Metasuchern mehrere Datenbanken auf einmal anzapfen kann, deckt er mit einer Anfrage einen größeren Bereich des Netzes ab, als er es mit nur einer Suchmaschine könnte.

Das Hauptproblem bei der Recherche im Internet besteht aber wie schon erwähnt nur selten darin, möglichst viele Treffer zu erhalten, sondern eine überschaubare Anzahl relevanter Ergebnisse zu finden. Daher eignen sich Metasucher nur für Recherchen in sehr speziellen Themengebiete oder für Suchen, die möglichst umfassend ausfallen sollen - zumal die Metasucher nur einen

kleinen Ausschnitt der Filtermöglichkeiten bieten, über die viele Volltextsuchmaschinen verfügen.

### 2.3.1. Kategorien von Suchmaschinen

- ***Weltweite Suchmaschinen***

⇒ Durchsuchen das gesamte Internet

[www.alltheweb.com](http://www.alltheweb.com)

[www.hotbot.com](http://www.hotbot.com)

[www.excite.com](http://www.excite.com) (de)

[www.yahoo.com](http://www.yahoo.com) (de)

[www.lycos.com](http://www.lycos.com) (de)

[www.infoseek.com](http://www.infoseek.com) (de)

[www.webcrawler.com](http://www.webcrawler.com)

- ***Deutsche Suchmaschinen***

⇒ Nur deutschsprachige Web-Seiten, bzw. bei Fireball werden Suchanfragen für nicht-deutschsprachige Seiten an Altavista weitergegeben.

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

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

[www.dino-online.de](http://www.dino-online.de)

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

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

- ***Meta-Suchmaschinen***

⇒ Fragen mehrere einzelne Suchmaschinen ab und liefern ein Gesamtergebnis zurück

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

[www.metacrawler.com](http://www.metacrawler.com)

[www.suchen.com](http://www.suchen.com)

### 2.4. Portale

Portale sind spezielle WebSites, die zu einem bestimmten Themenbereiche Adressen zur Verfügung stellen haben. Sie sind also eine Art Linksammlung.

z.B. [www.reiseplanung.de](http://www.reiseplanung.de)

Das heißt, eine Redaktion hat alle (wenn das überhaupt möglich ist) bekannte Seiten zu einem bestimmten Thema gehörenden Seiten gesammelt und geordnet.

Diese Form der Adressensammlung setzt sich in den letzten zwölf Monaten mehr und mehr durch.

Sie erlauben einen schnellen und umfassenden Überblick über die zu einem Thema gehörenden Adressen.

Themen können sein: Reiseplanungen, Übersetzungen, Finanzdienstleistungen und natürlich Suchmaschinen.

WWW.DOZENTEN-TEAM.DE

### 3. Informationen suchen und finden

*„Das Internet kann man sich vorstellen wie eine große Bibliothek – nachts, ohne Beleuchtung und mit allen Büchern auf dem Boden verteilt.“*

Anthony Rutkins

Das Internet ist eine Fundgrube für Informationen jeder Art. Das einzige Problem besteht darin, diese Information zu finden.

#### 3.1. Schnelles Auffinden aktueller Informationen

**Tagesaktuelle Informationen** bieten die Pressedienste, Tageszeitungen, TV-Sender und Onlinemagazine. Verwenden Sie einfach die Produkte, die Sie auch im „wahren Leben“ verwenden würden. In einem Katalog finden Sie die Adressen der Seitenanbieter mit aktuellen Meldungen. Suchen Sie ein bestimmtes Datum, einen Namen, oder einen Sachverhalt, dann nutzen Sie eine oder mehrere Volltextsuchmaschinen.

#### 3.2. Der richtige Einsatz von Suchmaschinen

##### 3.2.1. Die Stichwortsuche

**Volltextsuchmaschinen** arbeiten mit Stichworten, deren Bedeutung sie nicht verstehen, so dass Sie als Suchende möglichst stark eingrenzen sollten. In diesem Zusammenhang ist es sinnvoll, sich neben der Verknüpfung von Suchbegriffen auch über die Häufigkeit des Vorkommens Gedanken zu machen.

Daraus ergibt sich folgende **Hierarchie bei der Verwendung von Suchbegriffen**:

- **Einzelner, komplizierter, seltener Begriff.**
- **Spezieller Begriff zu einem speziellen Thema.**
- **Spezieller (Ober-)Begriff, der in mehreren unterschiedlichen Zusammenhängen auftauchen kann.**
- **Gebräuchlicher Begriff, der in mehreren unterschiedlichen Zusammenhängen auftauchen kann.**
- **Oberbegriff.**
- **Allgemeiner Oberbegriff, Kategorie.**
- **Häufig gebrauchter allgemeiner Begriff.**
- **Füllwort, Artikel, Pronomen**



### Zusammenfassung:

**Komplizierte Fachbegriffe** und ähnliches werden eine **überschaubare Menge an Treffern** liefern.

**Allgemeine Begriffe** oder gar Füllworte hingegen **liefern eine zu große Treffermenge**, deren Auswertung den Rahmen des Möglichen sprengt, da die Auswertung zuviel Zeit in Anspruch nimmt.

### 3.2.2. Wie fragt man richtig?

Das wichtigste bei jeder Internet-Recherche ist Fragetechnik. Wenn man in den Suchmaschinen nur die richtigen Suchbegriffe eingibt, erhält man schon mal ein gutes Ergebnis. Wenn man dann noch die **richtige Kombination** mit den **passenden Verknüpfungen** eingibt, erhält man mit Sicherheit (je nach Thema) eine Vielzahl von gewünschten Ergebnissen mit relativ wenig unerwünschtem Datenmüll. Vor Beginn einer Recherche einige Gedanken machen:

- Welche Informationen möchten Sie erhalten?
- Thematik, Institutionen, Organisationen, Unternehmen?
- Begriffe, Benennungen, Umschreibungen?
- Geographischer Bereich?
- Spezifische Informationen?

Bevor ein Surfer also vorschnell einen Recherchebegriff in eine der Suchmaschinen eintippt, sollte er sich genau überlegen, **wo er sucht**. Häufig führen gerade nicht die Suche mit einer der Volltextsuchmaschinen oder gar einer Metasuchmaschine am schnellsten zur Antwort, sondern die Recherche nach einem **Spezialangebot**, das die benötigte Information bereithält. Mittlerweile gibt es Hunderte von **Spezialdatenbanken (=Portale)**, die jedermann kostenlos offen stehen.

Darüber hinaus sollte der Suchende genau erwägen, über welche Informationen er bereits verfügt und **was genau er sucht**. Je spezieller er den Suchbegriff wählt, desto geringer ist die Wahrscheinlichkeit, dass dieser Begriff in einem anderen Zusammenhang vorkommt als in dem gewünschten. Das schöne Wort *Buttermilchproduktion* zum Beispiel ist dermaßen einem bestimmten Fachgebiet zuzuordnen, dass es kaum in einem falschen Zusammenhang vorkommen kann. In solchen Fällen ist es sinnvoll, eine Suchanfrage zu starten, die eine möglichst breite Datenbasis erfasst. Metasuchmaschinen sind hierfür besonders geeignet. Es kann bei einem speziellen Begriff vorkommen, dass selbst Metasuchmaschinen nur eine Handvoll Einträge finden.

Aber selbst Fachbegriffe kommen hin und wieder in unterschiedlichen Zusammenhängen vor. Wer feststellt, dass eine breit angelegte Metasuche zu viele Ergebnisse liefert, muss diese einschränken. Leider lassen **Metasuchmaschinen keine allzu komplizierten Verknüpfungen** von Suchbegriffen zu, so dass es notwendig werden kann, zu einer normalen Volltextsuchmaschine zu wechseln. Dort kann der Surfer dann die verschiedenen **Filtermöglichkeiten** einsetzen.

Hier ist eine Vorauswahl nach dem Sprachraum sinnvoll: Suchmaschinen mit internationalem Datenbestand müssen die Webadressen weltweit durchkämmen, deutschsprachige Suchdienste, die ihre Dokumente nach der verwendeten Sprache filtern und nur passende Seiten aufnehmen, können die Gesamtzahl an Einträgen für diesen etwas kleineren Suchraum leichter aktuell und weitgehend komplett halten.

Grundsätzlich sollte man **einige Tipps** beherzigen, wenn man im Internet schnell an die gewünschten Informationen herankommen will:

- **Groß- und Kleinschreibung** wird von manchen Suchmaschinen automatisch umgewandelt, es spielt dann also keine Rolle, ob Sie **AUTO**, **auto** oder **Auto** eingeben, das Ergebnis bleibt gleich.

**Nachteil:** Aber eben nicht immer! Deshalb sollten Sie vorsichtshalber bei der Recherche die vollständige Kleinschreibung praktizieren. Die reine Kleinschreibung veranlasst bei vielen Suchmaschinen, dass das Wort unabhängig von der Groß- und/oder Kleinschreibung gefunden wird. Wenn man auch nur einen Großbuchstaben verwendet, wird exakt nur diese Schreibweise des Wortes gesucht.

Also wird das Wort **auto** in einer Abfrage die Begriffe **auto**, **Auto**, **aUtO**, **AUTO** etc. in einem Dokument finden. Wird allerdings das Wort **Auto** eingegeben, wird auch nur nach genau dieser Schreibweise des Wortes **Auto** gesucht, nicht aber nach Varianten der Groß- und Kleinschreibung.

- Gesucht wird meistens auch nach **Wortbestandteilen**, d.h., die Eingabe von **mobil** führt dazu, dass u.a. auch nach **Automobil** und **Mobilfunk** gesucht wird.
- Jeder Suchbegriff wird meist als ein eigenständiges Wort interpretiert. Die Eingabe von **mobil club** führt z.B. zum Ergebnis **Automobil Club**.
- Falls Ihre Suchanfrage erfolglos war, versuchen Sie einen **kürzeren Begriff**. Geben Sie **Auto** statt **Automobil** ein, dann werden z.B. auch die Begriffe **Automobil Club**, **Automobil** und **Autohandel** gefunden.
- **Verwenden Sie klare Begriffe:** Man sollte möglichst präzise angeben, was man will. Unklare Begriffe, wie z.B. **berühmte Regisseure** verwirren auch die beste Suchmaschine, die dann viel Datenmüll präsentiert. Am besten gibt man mehrere verwandte Begriffe wie **Film**, **Entertainment**, **Regie** an, um die Trefferquote zu erhöhen. Ein Pluszeichen, z.B. **+Spielberg**, sorgt bei vielen Suchmaschinen dafür, dass nur Dokumente angezeigt werden, in denen dieser Begriff in jedem Fall auftaucht.
- **Unterschiedliche Schreibweisen:** Oft werden Namen und Begriffe in anderen Sprachen anders geschrieben, z.B. **Elisabeth** und **Elizabeth**. Manchmal schleicht sich auch ein Tippfehler (z.B. **Ellisabeth**) ein, oder eine eigenwillige Schreibweise (z.B. **Elisabet**) wird bevorzugt. Man sollte also unterschiedliche Schreibweisen ausprobieren, wenn man nicht direkt zum Ziel kommt.

- **Singular statt Plural:** Die meisten Suchmaschinen können nicht vom Plural auf den Singular schließen. Umgekehrt ist das meist kein Problem (z.B. Film statt Filme oder movie statt movies). Lassen Sie deshalb lieber den Singular suchen.
- **Joker benutzen:** Wenn man auf jeden Fall auch zusammengesetzte Begriffe und alle möglichen Endungen finden will, sollte man den Joker benutzen (\*, auch Wildcard genannt). Die Suche nach Film\* zum Beispiel bringt u.a. die Begriffe Filme, Filmemacher und Filmen.

**Achtung:** Das \*Zeichen ist ebenfalls recht nützlich, wenn man sich über die präzise Schreibweise nicht einig ist oder sie nicht kennt. So wird zum Beispiel mit der Eingabe a-lumi\*m nach der US-Schreibweise aluminum sowie der englischen Schreibweise alumin-um gesucht.

Eines sollte man aber beachten: Überlegen Sie sich den Einsatz des \*-Zeichens sehr gut! Wenn sie sing\* eingeben, um die Worte sing, singer, singers und singing zu finden, werden auch singe, single, singular, das französische Wort singulier und viele verwandte Wörter gefunden!

- **Verknüpfungen verwenden:** Um die Inhalte der zu findenden Dokumente möglichst präzise an sein Thema heranzuführen, sollte man mehrere Begriffe verknüpfen, die in einem Dokument vorhanden sein müssen. Die Boolesche Logik der Suchmaschinen arbeitet meist mit AND, BUT, OR und NOT. So wird die Suche Film AND Regisseur eher zu Steven Spielberg führen als der Suchbegriff berühmter Regisseur.
- **Ausdauer zeigen:** Die menschliche Kreativität kann so manches Mal den Computer und die Suchmaschinen überlisten. Deshalb sollte man hartnäckig bleiben und versuchen, die unterschiedlichsten Möglichkeiten einzugeben. Wechseln Sie auch schon mal die Suchmaschine, denn nicht jede Suchmaschine arbeitet nach demselben Prinzip.

### 3.2.3. Suchtechniken anwenden

Das Problem mit der Suche im Netz besteht inzwischen weniger darin, überhaupt einen Treffer zu finden, als vielmehr, viel zu viele irrelevante Fundstellen gelistet zu bekommen. Folglich bleibt keine andere Wahl, als sich mit den Möglichkeiten zu befassen, eine Suche einzugrenzen. Die Funktionen dazu sind je nach Suchdienst verschieden.

Als Quasi-Standard kann man aber die **booleschen Operatoren** benutzen, die mit wenigen Ausnahmen so gut wie jede Suchmaschine beherrscht.

Bei Verwendung der **UND-Verknüpfung** werden nur Fundstellen angezeigt, in denen alle mit UND verknüpften Suchbegriffe vorkommen. Das Pendant zum UND-Operator ist in internationalen Suchmaschinen das AND. Die Verknüpfung funktioniert auch über ein Pluszeichen „+“, das jedem einzelnen Begriff unmittelbar voranzustellen ist.

Bei Verwendung der **ODER-Verknüpfung** werden alle Fundstellen angezeigt, in denen mindestens einer der mit ODER verknüpften Suchwörter vorkommt. Bei englischsprachigen Suchmaschinen wird statt des ODER-Operator ein OR verwendet.

**Das Minuszeichen** „-“ schließt das Vorkommen des jeweiligen Suchbegriffs explizit aus. In der Treffermenge werden nur solche Dokumente angezeigt, in denen der Begriff nicht vorkommt. Auch dieses Zeichen muss unmittelbar vor dem betreffenden Wort stehen. Anstelle des Minuszeichens kann auch NICHT oder NOT vor ein Suchwort gesetzt werden.

**Trunkierungen** (=“Text\*“) ermöglichen es, nach Wortstämmen zu suchen, indem der Suchbegriff an passender Stelle einfach abgekürzt wird. Die meisten Maschinen verwenden hierzu den Stern „\*“: Die Suche nach Auto\* findet dann beispielsweise Automobil, Automotor, Autobahn etc. Genauso nützlich ist die Möglichkeit, mehrere Wörter zu einer Phrase zusammenzufassen. Hierzu werden meist Anführungszeichen verwendet: „New York“ findet genau diese Zeichenkette inklusive der Leerzeichen und nicht die einzelnen Wörter.

Durch die **Kombination von logischen Operatoren** – vor allem in Verbindung mit Klammern – können einfache bis sehr komplexe Suchanfragen formuliert werden. So lässt sich einerseits die Treffermenge auf ein überschaubares Maß reduzieren. Andererseits können auch unerwünschte Fundstellen ausgefiltert werden. Wenn Sie beispielweise nach einem weißen Auto suchen, aber auf keinen Fall ein Cabrio möchten, könnte Ihre Suchanfrage so aussehen: **(Auto UND weiß) NICHT Cabrio**.

Einen Spezialfall der Verknüpfung stellt die Phrasensuche dar, bei der auf zusammenhängende Begriffe (die wenigstens ein Leerzeichen enthalten) wie „Los Angeles“ oder „französische Küche“ geprüft wird. Phrasen werden wie im Beispiel entweder über **Anführungsstriche** oder eine entsprechende Auswahloption in der Suchmaske.

Die meisten Suchmaschinen und Kataloge bieten auch eine erweiterte Suche, Profisuche, Detailsuche an, um die Suche einzugrenzen, insbesondere hinsichtlich der Suchorte (z.B. nur deutschsprachige Websites), der Dokumentenart (z.B. Suche nur nach Bildern oder Texten). Ebenfalls ist bei den Suchmaschinen und Katalogen in der Regel eine Hilfe-Funktion vorhanden, in der man alle nötigen Informationen zur Einschränkung der Suche erhält. Nutzen Sie die jeweilige Hilfe-Funktion, um Ihre Suchanfragen für die jeweilige Suchmaschine, den Katalog korrekt zu formulieren.

### 3.2.3.1. Einfache Suchmethoden

+	+Microsoft +Outlook	Veröffentlichungen über Microsoft und Outlook
-	„ja wo laufen sie denn?“	Web-Seiten, die die Wortkombination „ja wo laufen sie denn?““ enthalten
„“	+„New York“ - Brooklyn	Web-Seiten zu New York, in denen das Wort Brooklyn nicht vorkommt
*	Bank*	Web-Seiten, in denen Worte vorkommen, die mit Bank
<b>AND</b>	BMW AND Motorrad	Web-Seiten über BMW und Motorräder
<b>OR</b>	Messen OR Ausstellungen	Veröffentlichungen über Messen oder Ausstellungen
<b>AND NOT</b>	Hamburg AND NOT Clubs	Web-Seiten über Hamburg, in denen nicht von Clubs die Rede ist
<b>NEAR</b>	Hannover NEAR Cebit	Web-Seiten, in denen der Begriff Cebit höchstens 10 Worte entfernt von dem Begriff Hannover steht.
<b>()</b>	„Los Angeles“ AND (Hollywood OR Malibu)	Web-Seiten über Los Angeles, in denen entweder von Hollywood oder von Malibu geschrieben wird

Bei der Bildung von Suchabfragen mit Hilfe logischer Operatoren müssen Sie die Reihenfolge der Auswertung beachten. Die höchste Priorität bei der Auswertung hat der **AND**-Operator, er wird zuerst berechnet, danach folgen der **AND NOT**- und der **OR**-Operator (vgl. Punkt- vor Strichrechnung bei den Grundrechenarten). Um unerwünschte Ergebnisse zu Vermeiden, können Sie die Reihenfolge der Auswertung durch Setzen von Klammern eindeutig festlegen.

### 3.2.3.2. Erweiterte Suchmethoden

Viele Suchmaschinen bieten komfortable weitere Einschränkungsmöglichkeiten an. Es lohnt sich, einmal einen Blick auf die **erweiterten Funktionen** zu werfen: „Advanced Search“, „Detail-suche“ oder „mehr Optionen“ sind Links, hinter denen sich eine ganze Reihe weiterer Konfigurationsmöglichkeiten verbergen. Ein wichtige Möglichkeit ist die Eingrenzung auf einen bestimmten Erstellungszeitraum; so lassen sieh Verweise auf veraltete Dokumente ausblenden.

Aber Vorsicht: Wenn man bei Suchmaschinen die Eingabefelder **START DATE** und **END DATE** ausfüllt, sollte man wissen, dass man damit für die Recherche das Eingabedatum des Dokuments in die Datenbank festlegt! Auch der gezielte Ausschluss von Domains kann sinnvoll sein - vor al-

lem dann, wenn viele Treffer auf einer einzelnen Website das Gesamtergebnis durcheinander bringen.

Die Ergebnisse einer erweiterten Suche werden oft ohne jegliche Gewichtung angezeigt. Bei einer präzisen Formulierung der Eingabe reicht das manchmal aus. Falls die Anzahl der gefundenen Dokumente aber zu umfangreich ist, sollte man sich eine Reihenfolge anzeigen lassen, wenn dies möglich ist. In diesem Fall muss man neben dem SEARCH FIELD auch ein RANKING FIELD (oder ähnlich) ausfüllen. Dort wird dann festgelegt, welches Wort oder welcher Begriff die Rangfolge der gefundenen Seiten bestimmen soll.

### *Erweiterte Suche mit Schlüsselwörtern*

<b>anchor:</b>	Suche nach Seiten, auf denen eine Textmarke oder ein Anker zu einem bestimmten Thema gesetzt wurde
<b>from:</b>	Suche nach Veröffentlichungen von bestimmten Personen
<b>host:</b>	Suche nach bestimmten Hosts- oder Servern
<b>image:</b>	Suche nach Bildern zu einem bestimmten Thema
<b>link:</b>	Suche nach Links, die auf eine bestimmte Web-Präsentation gesetzt sind
<b>newsgroup:</b>	Suche nach Newsgroups (Diskussionsforen) zu einem bestimmten Thema
<b>text:</b>	Suche nach einem Begriff, der im Text von Web-Seiten vorkommt
<b>title:</b>	Suche nach Veröffentlichungen mit einem bestimmten Titel
<b>url:</b>	Suche nach einer Internetadresse oder Bestandteilen einer Internetadresse
<b>domain:</b>	Suche nach Seiten aus einem bestimmten Land
<b>applet:</b>	Suche nach einem JavaApplet mit einem bestimmten Namen

*Beispiele*

<b>anchor:</b>	anchor:Euro	Sie möchten Seiten finden, bei denen <i>Euro</i> im Text des Hyperlinks steht
<b>from:</b>	from:*Henn	Sie suchen Veröffentlichungen von Personen mit Namen Henn. Der Vorname wird in der Suche nicht berücksichtigt.
<b>host:</b>	host:uni-kl	Sie können Seiten auf bestimmten Hosts oder Servern suchen. Im Beispiel suchen Sie alle Seiten, die auf dem Server der Universität Kaiserslautern liegen.
<b>image:</b>	image:europa.gif	Sie suchen Bilder zu einem bestimmten Thema. In der Regel haben Bilder in Web-Präsentationen die Endung *.gif oder *.jpg.
<b>link:</b>	link:herdt.de	Sie möchten erfahren, welche Firmen, Institutionen oder Privatpersonen einen direkten Verweis auf die Web-Präsentation von Herdt gesetzt haben.
<b>newsgroup:</b>	newsgroup:computer	Sie suchen eine Newsgroup (Diskussionsforum), die sich zum Thema Computer austauscht.
<b>text:</b>	text:Euro +text:Geldnoten + text:Einführung +text:2001	Sie suchen Veröffentlichungen, die an einer beliebigen Stelle der Web-Seite die Begriffe Euro, Geldnoten, Einführung und 2001 enthalten
<b>title:</b>	title:„Hamburger Fischmarkt“	Sie suchen alle Veröffentlichungen, bei deren Web-Seiten-Titel Hamburger Fischmarkt vorkommt
<b>url:</b>	url:ibm	Sie suchen Seiten, die ibm in der Adresse enthalten.
<b>domain:</b>	domain:ch	Sie suchen Seiten aus der Schweiz.
<b>applet:</b>	applet:money	Sie suchen ein JavaApplet mit einem bestimmten Namen.

### 3.3. Unterschiedliche Suchmaschinen und Ihre Bedienung

Name	Suchraum:		Logische Verknüpfung				Weitere Funktionen	
	Land. Sprache	Suche in:	UND	ODER	NICHT	NEAR	PHRASEN	Trunkierung
Acoon	Deutschsprachige Seiten unter allen Domains	WWW	+a +b	a b	a -b	nein	nein	nein
Aladin	Deutschsprachige Seiten unter allen Domains	WWW	a AND b	a OR b	nein	nein	„Boris Becker“	Suchwo*
Alta Vista	International, verschiedene Sprachen wählbar, Spracherkennung	WWW, News	+a +b	a b	a -b	nein	„Boris Becker“	Suchwo*
			a AND b	a OR b	a AND NOT b	a NEAR b (10 Worte)		
Crawler.de	Deutschsprachige Seiten	WWW	Automatisch	-	nein	nein	nein	Nein
Eule	Deutschsprachige Seiten .de	WWW	+a +b	a b	a -b	nein	nein	Nein
Excite.de	International, deutschsprachige und europäische Domains	WWW, News	+a +b a AND b	a OR b	a AND NOT b	nein	„Boris Becker“	Nein
Fireball	Deutschsprachige Seiten, Spracherkennung	WWW	+a +b	a b l	nein	nein	„Boris Becker“	Suchwo*
			a & b a AND b	a ! b a OR b	a   b a AND NOT b	a ~ b a NEAR b		
Hot Bot	International	WWW, News, Yellow pages, E-Mail Adr. Shareware	a +b	a b	a -b	nein	„Boris Becker“	nein
			a AND b a & b	a OR b a   b	a NOT b a ! b	nein		
Infoseek	International	WWW, News	+a +b	a b	a -b	nein	„Boris Becker“	Automatisch
Speedfind	Deutschsprachige Domains .de .at	WWW, E-Mail Adr	+a +b	a ,b	a -b	a NEAR b(10 Worte)	„Boris Becker“	Suchwo* Suchwo?t
Lycos.de	Deutschsprachige Domains .de .at .ch .li, oder international	WWW, News, E-Mail Adr	Auswahl- button "alle Wörter"	Auswahl- button "ein Wort"	nein	nein	Auswahl- but- ton "genauer Text"	Automatisch
			a AND b	a OR b	a NOT b	a NEAR b	über Drop Down Menü	

Phrase=so gekennzeichnete Ausdrücke behandeln die Suchhilfen wie ein Wort  
 Trunkierung=Einsatz eines Platzhalters(Wildcard) für eine beliebige Anzahl von Zeichen; nützlich, um z. B. verschiedene Schreibweisen zu definieren: Del\*in findet sowohl Delfin wie auch Delphin.