

# Internet und Suchmaschinen

Erstellt von Penn Jack Wang , 3. Fachsemester Informatik

---

## Überblick

- 1.0 Einführung
  - 2.0 Suchwerkzeuge
    - 2.1 Kataloge
    - 2.2 Suchmaschinen
    - 2.3 Meta-Suchmaschinen
    - 2.4 Alternative Suchwerkzeuge
  - 3.0 Zukunft des Suchens
  - 4.0 Internet Historie
  - 5.0 Glossar
  - 6.0 Quellen
- 

## 1.0 Einführung

Kann man sich das sinnvolle Arbeiten im Internet ohne Suchmaschine vorstellen? Wahrscheinlich kaum. Vor zehn Jahren hingegen gab es noch nicht einmal das Wort 'Suchmaschine' in der deutschen Sprache, und auch heute sucht man's im Duden oder Brockhaus vergeblich. In der Realität des Internet haben die Suchmaschinen jedoch einen dramatischen Aufschwung erlebt.

Das Internet enthält eine gigantische Informationsmenge, die größte, die der Mensch jemals geschaffen hat. Gerade deshalb aber ist das Auffinden brauchbarer Informationen eine nicht ganz einfache Aufgabe. Stellen Sie sich die Bücher einer Bibliothek auf einem Haufen vor - kaum jemand wäre in der Lage, in akzeptabler Zeit das gewünschte Buch herauszusuchen. Jetzt stellen Sie sich alle Bibliotheken der Welt auf einem Haufen vor. Ungefähr so sieht es im Internet aus.

Ohne Suchmaschinen wären Rechercheaufgaben im Web von vornherein zum Scheitern verurteilt. Aber auch mit diesen unentbehrlichen Helfern ist der Erfolg nicht garantiert. Suchmaschinen haben so ihre Tücken: entweder ist das Gesuchte gerade nicht erfaßt oder aber der Suchende wird umgekehrt von der Menge des Gefundenen 'erschlagen'.

Die Internetsuchmaschinen haben ständig gegen den Zuwachs des Internet zu kämpfen. Denn das WWW umfasst ca. 5 Milliarden Seiten, davon kommen täglich 7 Millionen hinzu.

---

## 2.0 Suchwerkzeuge

Es gibt mehrere Suchwerkzeugarten, hier werden nur die üblichen beschrieben, nämlich Kataloge, Suchmaschinen und Metasuchmaschinen.

### 2.1 Kataloge

Bei einem Katalog handelt es sich um ein von Menschenhand zusammengestelltes Suchangebot. Dabei ist es in Kategorien eingeteilt und sehr übersichtlich gestaltet. Meistens verbirgt sich hinter einen Katalog eine nach alphabetisch und Kriterien geordnete Liste. Der Vorteil eines Katalog ist, daß die Seiten meist recht gute Qualität aufweisen, da sie von den Redakteure begutachtet werden. Ausserdem sind die Themen sehr gut eingeteilt. Der Nachteil bei Katalogen ist, daß sie nur einen kleinen Teil des WWW einnehmen und nicht mit den Wachstum des WWW mithalten können.

Beispiele einiger Kataloge: Yahoo!, Universität Hamburg Seite,...!

### 2.2 Suchmaschinen

Suchmaschinen gegenüber Katalogen sind komplexer aufgebaut. Suchanfragen werden durch komplexe Suchalgorithmen behandelt. Danach benutzt die Suchmaschine sogenannte Roboter, die die Webseiten durchforsten und Links folgen. Die Suchergebnisse werden dann in Datenbanken angeordnet. Um an einen erfolgreichen Ergebnis zu kommen, müssen schon boolesche Suchanfragen gestellt werden. Booleschen Ausdrücken sind zum Beispiel: "OR" und "AND"! Die Gewichtung von Suchergebnissen ist von Suchmaschine zu Suchmaschinen verschieden (Suchranking). Das Suchranking beschränkt sich entweder auf die Anzahl der Besucher der Webseite oder der Wichtigkeit der Links. Die Vorteile der Suchmaschinen können positiv als auch negativ angesehen werden. Zum einen erfasst die Suchmaschinen mehr Webseiten als Kataloge, dennoch sind es gegenüber des gesamten WWW nur 16 Prozent. Man findet zwar meistens den eingegebenen Suchbegriff, dennoch wird man von der Masse der Suchergebnisse erschlagen. Was zusätzlich zur den ständigen WWW-Wachstum kommt, ist das viele Inhalte der Webseiten oft veraltet sind oder sie führen zu toten Links.

Beispiel einiger Suchmaschinen: Altavista, Google, Lycos,...! Inzwischen gibt es über 1000 Suchmaschinen!

### 2.3 Metasuchmaschinen

Falls es den Benutzer immernoch nicht zu einen erfolgreichen Ergebnis gelungen ist, bieten Metasuchmaschinen das parallele Suchen an. Das bedeutet, daß sie mehrere herkömmliche

Suchmaschinen gleichzeitig durchsuchen und dann ein zusammengestelltes Suchergebnis ausgeben. Dabei muss jede Metasuchmaschine 6 von den folgenden 7 Kriterien beinhalten:

1. **Parallele Suche**(keine All-in-one-Formulare): Die Metasuchmaschine muß wirklich parallel suchen.
2. **Ergebnis-Merging**: Die Ergebnisse müssen zusammengeführt und in einem einheitlichen Format dargestellt werden.
3. **Doubletten-Eliminierung**: Doppelte Fundstellen müssen erkannt und gekennzeichnet werden.
4. **Mindestens AND - und OR-Operatoren**: Für logische Operationen müssen mindestens die Operatoren AND und OR zur Verfügung stehen.
5. **Kein Informationsverlust**: Wenn ein Suchdienst eine Kurzbeschreibung der Fundstelle liefert, dann muß diese übernommen werden.
6. **Search Engine Hiding**: Die spezifischen Eigenschaften der unter der Meta-Maschine liegenden Suchdienste dürfen für die Bedienung keine Rolle spielen, der Anwender muß nichts darüber wissen müssen
7. **Vollständige Suche**: Die Metasuchmaschine sollte in der Lage sein, so lange zu suchen, bis irgendeine der darunterliegenden Suchmaschine noch Treffer liefert.

Ein Nachteil den Metasuchmaschinen haben , ist daß sie auf komplexe Suchanfragen streiken.

Beispiele von Metasuchmaschinen: Metacrawler!

## 2.4 Alternative Suchwerkzeuge

Folgende Suchwerkzeuge habe ich hier als alternative Suchwerkzeuge angegeben , da man ihnen Suchanfragen geben kann und Suchergebnisse erzielen kann.

- **Newsgroup**:Bei der Newsgroup handelt es sich um Diskussionsforen , wo man gezielten Fragen stellen kann , die dann von anderen Benutzern beantwortet werden. Dabei geht es um Informationsaustausch, der nur begrenzt auf das Wissen des Benutzers beschränkt ist.
- **Offline Suchmaschinen**:Dabei handelt es sich um mehr eine Metasuchmaschine, die man bei Offlinebetrieb Anfragen stellen kann, die dann beim Ausführen sich ins Internet wählt und mehrere Suchmaschinen gleichzeitig durchsucht.  
Beispiel:iSeeker!
- **Alexa**:Dieses Programm, wenn installiert, zeigt alle verwandte Links der besuchten Seite auf. Bei mehr Informationen die Alexa Webseite liefert mehr!

---

## 3.0 Zukunft des Suchens

Die Zukunft des Suchens sieht folgendermaßen aus:

- **Gesichtserkennung mit Bildersuchmaschinen:** Dieses Verfahren wird von der Polizei schon seit Jahren benutzt um vermisste Personen im Internet aufzuspüren. Dabei merkt sich die Suchmaschine Merkmale des Gesichts (Bild) und speichert sie in Muster ab. Ein Programm , das nach diesem Verfahren arbeitet, ist der **Duplicate Image Finder** von Runningman Software
  - **Musiksuchmaschinen:** Damit sind nicht die Suchmaschinen gemeint , die nach MP3s suchen wie Napster oder Audiogalaxy , sondern Suchmaschinen, die besondere Soundstrukturen aus Musikstücken herauserkennen können und somit Musikpiraten auf die Schliche kommen.
  - **Suchmaschinen die auch sinnverwandte Themen bereitstellen:** Dabei handelt sich es Suchmaschinen die zum Beispiel bei dem Wort: Krieg; Themen wie Terror oder Frieden gleichzeitig ausgeben.
  - **Intelligente Suchmaschinen, die selbst handeln:** Suchmaschinen , die auf das Benutzerprofil zugeschnitten sind
- 

## 4.0 Internet Historie

- 1957: erste Entwicklung eines Netzwerkes zur militärischen Zweckes, Gründung der ARPA
  - 1969: ARPA nahm das erste paketorientiert arbeitende Netz in Betrieb (ARPANET)
  - 1977: wurde das ARPANET mit einem Satelliten- und einem Funknetz sowie dem von XEROX-PARC entwickelten Ethernet über TCP/IP verbunden - das Internet war geboren!
  - 1983: ARPANET umgestellt nach TCP/IP, andere Netze schlossen sich dem Internet zu (Usernet, Bitnet)
  - 1989: ARPANET vom Militär aufgelöst
  - 1990: Am Genfer Hochenergieforschungszentrum CERN überlegen Wissenschaftler, wie man für die zivile Nutzung Dokumente weltweit abrufen und Grafiken einbinden kann. Gründung des HTTP, WWW
  - 1993: Entwicklung des Browser Mosaic, Netscape, erste Suchmaschine "Wanderer"
- 

## 5.0 Glossar und Abkürzungen

**ARPA:** Advance Research Project Agency

**CERN:** Centre Européen de Recherches Nucléaires (Europäisches Kernforschungszentrum in Genf, Schweiz).

**HTTP:** Hypertext Transfer Protocol (Hypertext-Übertragungsprotokoll). HTTP ist ein Standard zur Übermittlung von HTML-Seiten im Internet.

**Roboter:** (siehe andere Referate)

**TCP/IP:** Transmission Control Protocol/Internet Protocol! Die Daten werden in handliche Pakete zerlegt und über das Netz geschickt. Den Weg finden sie mit der individuellen IP (Internet Protocol)-Adresse des Zielcomputers. Dort setzt TCP die Daten wieder automatisch zusammen.

**WWW:** World Wide Web

---

## 6.0 Quellen

- Online Internet Kurs
  - Verteilte System Fachbereich Informatik Uni Hamburg
  - Einführung ins Internet
-