


Lehrer: Spahr Marcel	SCHULE LAUPEN 	Lehrgang: NMM Informatik: Das Internet
Name:	SUCHEN IM WWW	Datum:

Bilder und Textpassagen stammen mehrheitlich aus dem Buch

[**Informationsbeschaffung im Internet Grundlegende Konzepte verstehen und umsetzen**](#)

Von Werner Hartmann, Michael Näf, Peter Schäuble

Illustrationen: François Chalet

1. Lernziele

Nachdem du dieses Kapitel durchgearbeitet hast,

- kennst du den Unterschied zwischen einer Suchmaschine und einem Katalog und weisst, wie die beiden Werkzeuge funktionieren.
- Kennst du verschiedene Möglichkeiten, wie man im WWW etwas gezielt sucht und nicht in der Informationsflut ertrinkt.
- Kannst du die Qualität und Glaubwürdigkeit der Daten und Informationen auf dem Internet beurteilen.
- Kannst du eine gefundene Seite als Bookmark abspeichern und mit Hilfe der Bookmarks wieder aufrufen.
- Kannst du die gefundenen Seiten auf deinem Computer abspeichern sowie Bilder und Text in dein Textverarbeitungsprogramm übernehmen.

2. Suchmaschine (auch Spider oder Crawler genannt)

Suchmaschinen sind Computer, welche für dich das WWW nach Stichworten durchsuchen. Eine Suchmaschine ist ein Dienst auf dem Internet mit dem man Teile des Internets durchsuchen kann. Eine Suchmaschine kann mit einem Index oder einem Schlagwortverzeichnis in einem Buch verglichen werden.

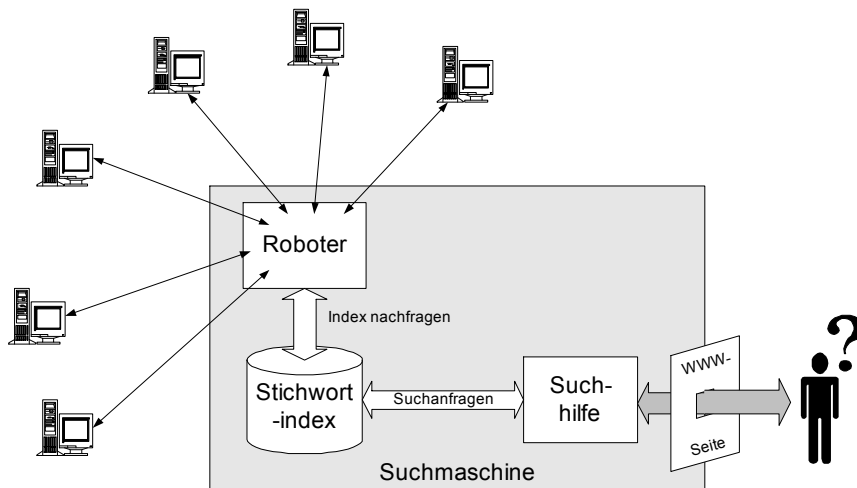
Für den Benutzer präsentiert sich eine Suchmaschine auf dem WWW wie jede andere WWW-Seite. In einem Formular können Suchkriterien eingegeben werden. Daneben kann die Suche auf bestimmte Objekte (Bilder, Tabellen), bestimmte Adressen (z.B. nur Schweizer Webserver) oder auf einen bestimmten Datumsbereich eingeschränkt werden. Die Suchmaschine liefert dem Benutzer eine Liste mit WWW-Adressen, die seinen Kriterien am besten entsprechen.



Auf dem Internet existieren unzählige Suchmaschinen. Gewisse versuchen das ganze WWW abzudecken (z.B. <http://www.altavista.com>), andere suchen nur in einem bestimmten Teil des WWW (z.B. kann man unter der Adresse <http://www.google.ch> nur Resultate aus der Schweiz anzeigen lassen).

Wie funktioniert eine Suchmaschine?

Eine Suchmaschine durchsucht das Internet nicht dann, wenn der Benutzer eine Suchanfrage stellt. Bei den Milliarden von Dokumenten auf dem Internet würde das viel zu lange dauern! Eine Suchmaschine besteht aus den drei Teilen Roboter, Index und Benutzer-Schnittstelle:



- Die **Benutzerschnittstelle** (Interface) ist das Formular und die Webseite, die der Benutzer sieht.
- Der **Stichwortindex** ist eine grosse Datenbank.
- Der **Suchroboter** durchstöbert dauernd das WWW und indexiert dabei gefundene Seiten, d.h. er nimmt die Seiten in einem Stichwortverzeichnis auf.

Ein Suchroboter beginnt seine Wanderung bei ein paar fest vorgegebenen Web-Seiten und folgt dann allen Links, die er auf diesen Seiten findet. So pflügt er sich durch immer grössere Bereiche des Webs. Er nimmt dabei alle Seiten, die er auf seinem Weg findet in seinen Index auf. Je nach Suchmaschine indexiert er nur Titel und die ersten paar Zeilen jedes Dokuments oder gleich das ganze Dokument. Grosse Suchmaschinen erfassen Unmengen von Web-Seiten. 2001 umfasste der Index der Suchmaschine Altavista gegen 520 Millionen Web-Seiten und lagerte in ihrer Datenbank über 1000 Gigabyte Daten (1000 GByte entsprechen rund 10^{12} Zeichen).

Nachdem du nun die Funktionsweise einer Suchmaschine kennst, sind auch die folgenden Punkte klar:

- Suchmaschinen finden nur WWW-Seiten, auf die ein Link zeigt. Ein einzelnes Dokument oder eine Gruppe von Dokumenten, auf die niemand hinweist, wird der Suchroboter nie finden. Aus diesem Grund besitzen Suchmaschinen meist eine Möglichkeit, neue Seiten anzumelden.
- Suchmaschinen finden selten topaktuelle Inhalte. Ein Suchroboter benötigt einige Zeit, bis er das WWW durchstöbert hat und mit der Suche wieder von vorne beginnt (z.B. 1 Monat). Es kann auch passieren, dass die Suchmaschine auf eine Seite hinweist, die der Suchanfrage gar nicht entspricht. Dann hat der Autor der Seite den Inhalt seit dem letzten Besuch des Suchroboters geändert.

3. Katalog auch Verzeichnis oder subject tree genannt

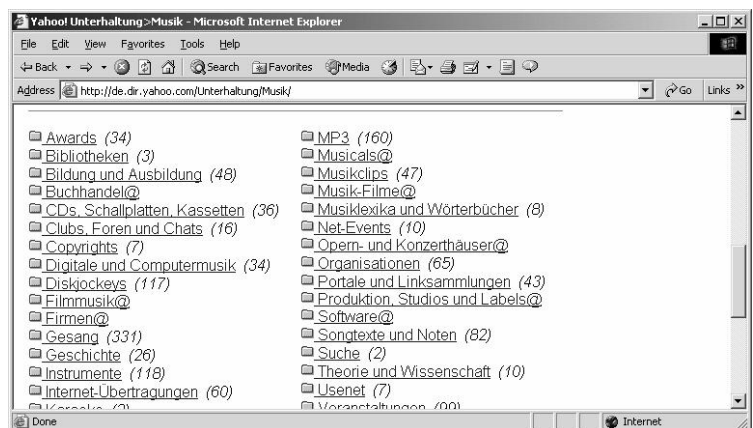
Ein Katalog ist eine nach Themen geordnete Sammlung von Links auf dem WWW. Kataloge werden im Gegensatz zu Suchmaschinen nicht automatisch erstellt. Ein oder mehrere Redaktoren versuchen wichtige WWW-Seiten in ihren Katalog aufzunehmen. Diese Seiten findest du entweder durch Surfen auf dem Web oder indem du von Autoren von Seiten oder von anderen Surfern darauf aufmerksam gemacht wirst.

Es existieren weltweite Kataloge, die versuchen, das ganze WWW abzudecken (Zum Beispiel <http://www.yahoo.com>). Daneben gibt es viele spezialisierte Kataloge, die sich auf ein oder mehrere Themen beschränken. (<http://www.music.ch> bietet z.B. nur Links zur Musikszene in der Schweiz).

Eigentlich ist beinahe jede Seite im WWW ein Kleinstkatalog, denn sie bietet Links zu einem oder mehreren verwandten Themen. Man spricht aber erst von einem Katalog, wenn sich die Seite als Einstieg ins WWW oder in ein spezielles Thema eignet und die Seite regelmässig aktualisiert wird.

Im grössten weltweiten Katalog <http://www.yahoo.de> findet man unter „Musik“ viele Unterkategorien zum Thema Musik.

<http://www.music.ch> ist ein Schweizer Katalog, der sich ausschliesslich mit Musik beschäftigt.



Unter der Endung .ch findet man meistens Angebote, die auf die Schweiz zugeschnitten sind, z.B. Schweizer Firmen oder die Schweizer Kantone. Internationale Firmen haben die

Endung .com in der Adresse. Non-Profit-Organisationen haben oft die Endung .org, z.B. www.greenpeace.org.

Rangierungsprinzipien: So liefern Suchmaschinen das Gesuchte!

Tja, Suchen will gelernt sein! Beim Computer sind es manchmal kleine Details, die viel ausmachen. In der folgenden Liste stehen 4 goldene Regeln, damit du die gesuchten Seiten leichter findest:

- **Goldene Regel 1:** Gib möglichst viele Suchbegriffe ein.
Mit dem Begriff „Gotthard“ alleine findet die Suchmaschine 18000 Seiten. Erst zusammen mit „concerts“, „official“ und „webSite“ hat die Suchmaschine die gewünschte Seite als relevant erkannt und zuoberst aufgeführt.
Tipp: Formuliere immer präzise Anfragen!
- **Goldene Regel 2:** Je häufiger ein Suchbegriff in einem Dokument vorkommt, desto wahrscheinlicher ist das Dokument relevant.
Konsequenz: Überlege, welche Begriffe im Dokument häufig auftauchen!
Beispiel: Wenn du einen neuen Bildschirm kaufen willst, so ist „resolution“ ein guter Suchbegriff.
- **Goldene Regel 3:** Dokumente, die seltene Suchbegriffe enthalten, sind mit einer höheren Wahrscheinlichkeit relevant als Dokumente, die häufige Suchbegriffe enthalten.
Konsequenz: Verwende spezifische Suchbegriffe!
Beispiel: Du suchst das Resultat des Schweizer Fussball-Cups von 1978. „Cupfinal Wankdorf“ ist eine genauere Anfrage als „Resultat Fussball-Cup“.
- **Goldene Regel 4:** Je näher die Suchbegriffe beieinander liegen, desto relevanter das Dokument.

Daneben gibt es je nach Suchmaschine noch weitere Regeln, wie die Dokumente rangiert werden.

Katalog

- 'Inhaltsverzeichnis' des WWW
- Thematisch geordnet
- Von Menschen erstellt
- Grosse Anzahl Links
- Gewisser „Qualitätsfilter“

Suchmaschine

- 'Index' oder Schlagwort-Verzeichnis des WWW
- Lexikalisch geordnet
- Automatisch erstellt
- Unbewältigbare Anzahl Links
- Ungefiltertes Angebot

- Geeignet als Überblick über ein Thema
- „Was finde ich im Gebiet ...?“
- Geeignet für konkrete Anfragen (Person, Produkt, Stichwort)
- „Ich weiss, dass es irgendwo...“

Erweiterte Möglichkeiten von Suchmaschinen

Viele Suchmaschinen bieten noch weitere Möglichkeiten um die Suche zu beeinflussen. Bei Google z.B. gibt es den Standard-Suchmodus, den du bereits kennst. Daneben existiert auch eine sogenannte 'Erweiterte Suche', eine Suchmaske für Fortgeschrittene. Um damit umzugehen, muss man sich etwas einarbeiten, dafür sind aber die Suchergebnisse einiges präziser.

Das Eingabeformular sieht etwas komplizierter aus als bei der Standard-Suche: In diesem Formular kann ich jetzt nicht nur einzelne Suchbegriffe eingeben. Ich versuche mal, Informationen zu Patent Ochsner zu finden:

Auf deutsch übersetzt heisst diese Frage in etwa:

'Ich suche alle deutschen Seiten auf dem WWW, die in den letzten 3 Monaten publiziert worden sind und auf denen der Ausdruck „Patent Ochsner“ steht. Liefere keine Seiten, in denen die Begriffe Abfall, Container und Behälter vorkommen. Beachte nur Domains, die mit .ch enden. Stelle das Resultat als Seiten mit je 10 Links dar.



Suchbegriff	Suchvorgang
mona lisa	Findet Dokumente, die „mona“ und „lisa“ in Gross- oder Kleinschreibung enthalten. Die gefundenen Ergebnisse werden von Google ausgegeben und eingestuft. Dabei werden zuerst die Dokumente aufgeführt, in denen beide Begriffe in unmittelbarer Nähe zueinander vorkommen.
mona -lisa	Findet Dokumente, die „mona“ in Gross- oder Kleinschreibung, aber nicht „lisa“ enthalten.
"mona lisa"	Findet Dokumente, in denen die beiden Wörter unmittelbar nebeneinander stehen. Wenn du mehrere Wörter durch Anführungszeichen bezeichnest, kennzeichnest du sie als eine Wortgruppe oder „Phrase“, dadurch weiss Google, dass du nur an Dokumenten interessiert bist, in diese Wörter in genau dieser Gruppierung auftreten.
"mona lisa" -louvre	Findet Dokumente, die „Mona Lisa“, aber nicht das Wort „Louvre“ enthalten.
mona lisa +.com	Findet Dokumente welche die Begriffe „mona“ und „lisa“ und „.com“

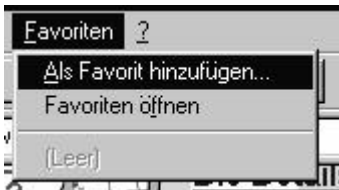
enthalten. Ohne das Pluszeichen würde der Begriff „.com“ nicht als Suchbegriff einbezogen. Bei Zahlen, einzelnen Buchstaben und bestimmten Wörtern wie "http" und ".com" handelt es sich um sogenannte Stop-Wörter, die von Google bei der Suche ignoriert werden, weil diese Begriffe nur bei der Einschränkung der Suche helfen, gleichzeitig aber die Geschwindigkeit bedeutend verlangsamen.

Benutze das "+"-Zeichen, um Stop-Wörter in Ihre Suche einzuschränken.

4. Search-Engines und Verzeichnisse

Google	www.google.ch oder www.google.de
Altavista	www.altavista.ch oder www.altavista.de
Yahoo!	www.yahoo.ch oder www.yahoo.de
Lycos	www.lycos.ch oder www.lycos.de
Webcrawler	www.webcrawler.de
Search	www.search.ch
Bluwin Search	www.sear.ch
Megasearch	(Webseite mit Links auf die wichtigsten Suchmaschinen) www.megasearch.ch

5. Bookmarks (auch Buchzeichen oder Favoriten)



Was machst du nun, wenn du deine absolute Lieblingsseite gefunden haben? Du kannst dir die Adresse dieser Seite aufschreiben. Wenn man immer am gleichen Computer arbeitet, kann das der Computer erledigen. Man muss es ihm nur sagen. Das Prinzip nennt sich Bookmarks, Favoriten oder auf deutsch: Buchzeichen.

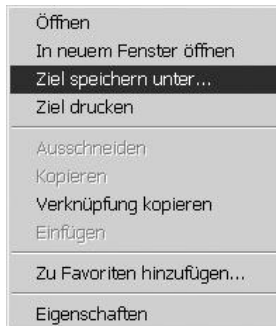
Bookmarks sind Links, die ein Browser auf Anweisung des Benutzers gespeichert hat. Bookmark ist das englische Wort für Buchzeichen und Bookmarks sind auch die Buchzeichen des WWW: Mit Hilfe von Bookmarks kann man Links auf Seiten, die man später wieder besuchen möchte, bequem abspeichern.

Wichtig: Es wird nur der Link, nicht aber die WWW-Seite auf dem Computer des Benutzers gespeichert.

Mit dem Menüpunkt „Add Bookmark“, merkt sich der Computer die Adresse der aktuellen Seite und trägt ihren Titel in die Bookmark-Sammlung ein. Wenn man später auf diese Seite zurückkommen möchte, so wählt man einfach aus der Bookmark-Sammlung (z.B. im Netscape Navigator im Menu 'Bookmarks') den entsprechenden Titel.

Je nach verwendetem Browser ist die Bookmark-Sammlung anders organisiert (beim Internet Explorer heissen die Bookmarks Favoriten). Das Prinzip ist aber immer dasselbe.

6. Das Gefundene speichern und weiterbearbeiten



Wenn du mit der rechten Maustaste auf ein Bild klickst, dann erscheint ein Menu. Wenn Du „Ziel speichern unter...“ wählst, dann kannst du einen Dateinamen angeben, unter dem das Bild auf Deiner Festplatte gespeichert wird. Dann kannst du es später, wenn du offline bist, in aller Ruhe bearbeiten oder ausdrucken. Das Gleiche kannst du auch bei Links machen. Dann wird das Dokument, auf das der Link zeigt, abgespeichert (z.B. eine mp3-Datei).“

Bei Computern mit nur einer Maustaste erscheint das Menu, wenn man die Maustaste lange genug gedrückt hält.

Wenn du nur den Text einer Webseite weiterverwenden willst, dann ist es oft einfacher, wenn du den gewünschten Text mit der Maus markierst und ihn dann in ein Textverarbeitungsprogramm kopierst, z.B. nach Word.



Du kannst auch die ganze Webseite auf deiner Harddisk abspeichern, wenn du im Datei-Menu auf „Speichern unter“ klicken. Das ist nützlich, wenn du einen langen Text in aller Ruhe lesen möchtest. Du kannst ihn natürlich auch ausdrucken.

7. Übungen

Aufgabe 1

2001 umfasste der Index der Suchmaschine Altavista gegen 520 Millionen Web-Seiten und lagerte in ihrer Datenbank über 1000 Gigabyte Daten (1000 GByte entsprechen rund 10^{12} Zeichen).

Versuche dir eine Vorstellung dieser Zahlen zu machen, indem du einen Vergleich machst in der Art von:

- Wenn ich 520 Millionen Papierseiten lesen müsste, so bräuchte ich dafür...
- Wenn man diese 1000 GByte ausdrucken würde, so würde der Stapel ...

Finde selbst einen originellen Vergleich! Folgende Zahlen können dir dabei helfen:

- Eine A4-Seite ist ca. 20 cm · 30 cm gross und wiegt ca 5g.
- Auf eine A4-Seite passen ca 5'000 Zeichen in Schreibmaschinenschrift.
- Ein Tintenstrahldrucker druckt etwa 3 Seiten pro Minute.
- Ein Eisenbahnwagen fasst ca. 40 Tonnen Altpapier und ist 14 m lang.

Überlege dir bei den folgenden Fragestellungen, ob du einen Katalog oder eine Suchmaschine verwenden würdest. Begründe deine Wahl in ein paar Worten. Falls du eine Suchmaschine verwenden würdest, notiere passende Stichworte!

Aufgabe 2

- Für einen Geschichtsvortrag suchst du die amerikanische Verfassung.
- Du suchst für die Sommerferien ein Hotel in der Nähe deines Traumziels.
- Du machst Ferien in Paris. Welche Museen sollte man gesehen haben und wie sind die Öffnungszeiten?
- Nach dem Museumsbesuch möchtest du in eine Disco. Wohin geht man?
- Du möchtest ein Funktelefon kaufen. Was ist in der Schweiz erhältlich?

8. Lösungen

Aufgabe 1

Mögliche Lösungen sind:

Angenommen, ich lese eine Textseite in 5 Minuten, dann brauche ich 5 520'000'000 Minuten. Diese 2,6 Milliarden Minuten entsprechen 43 Millionen Stunden, etwa 1,8 Millionen Tagen, also knapp 5000 Jahren. Das liest Altavista in ein paar Tagen!

Ich möchte die Datenbank von Altavista ausdrucken: $10^{12} / 5'000$ ergibt ungefähr $2 \cdot 10^8$ Seiten. Eine Seite wiegt 5 Gramm. Also wiegt der gesamte Stapel 10^9 g. Das entspricht 1 Million kg oder 1000 Tonnen. Der Güterzug wäre also 350m lang! 1999 umfasste das Bundesarchiv in Bern ebenfalls $2 \cdot 10^8$ Seiten. In Bundesordnern abgelegt benötigten diese über 40 km Platz. Wie hoch wäre der Turm, wenn man alle Seiten aufeinanderstapeln würde?

Aufgabe 2

Mögliche Stichworte und ihre Begründungen sind:

- Für einen Geschichtsvortrag suchst du die amerikanische Verfassung.
Die amerikanische Verfassung ist ein offizielles Dokument, das sicher auf den WWW-Seiten der amerikanischen Regierung zu finden ist. Ein Katalog hilft dir, rasch die

Seiten der US-Regierung zu finden. So bist du sicher, dass du nicht in Seiten ertrinkst, die nur auf die amerikanische Verfassung Bezug nehmen.

- **Du suchst für die Sommerferien ein Hotel in der Nähe deines Traumziels**
Da ist ein Katalog angezeigt! Entweder du beginnst mit der Stadt oder der Region in einem allgemeinen Katalog und findest dann Verweise auf Übernachtungsmöglichkeiten. Mit einer Suchmaschine und dem Begriff „Hotel“ neben deinem Traumziel findest du tausende von Reisebeschreibungen von Privatpersonen in der Art „In Venedig waren wir in einem wunderschönen Hotel auf dem Lido“ ...
- **Du machst Ferien in Paris. Welche Museen sollte man gesehen haben und wie sind die Öffnungszeiten ?**
Ein klarer Fall für den Katalog! Museen sind sicher in Katalogen zu finden. Einen guten Katalog zu den Pariser Museen findest du sicher auf der offiziellen Webseite zu Paris. Wie heisst die wohl? Kleiner Tipp: Das Länderkürzel von Frankreich ist fr.
- **Nach dem Museumsbesuch möchtest du in eine Disco. Wohin geht man ?**
Auch hier ist eine Suche via Katalog denkbar. Die offizielle Webseite zu Paris hat natürlich auch dazu einen Katalog. Aber wer wird bei einer Disco schon auf einer offiziellen Seite suchen... Versuche es mit einer Suchanfrage bei einer Suchmaschine.
- **Du möchtest ein Funktelefon kaufen. Was ist in der Schweiz erhältlich?**
Benutze eine Schweizer Suchmaschine. So erhältst du nur Modelle, die auch in der Schweiz erhältlich sind. Für technische Details musst du dann eventuell auf die Homepage des entsprechenden Herstellers gehen. Vielleicht errätst du auch die WWW-Adresse deines Elektronik-Händlers oder versuchst es direkt bei der Swisscom (<http://www.swisscom.com>).