

Werkstatt

„Suchen und Finden im Internet“

Thema

In der Werkstatt „Suchen und Finden im Internet“ lernst du, wie man Recherchen im Internet zielgerichtet und erfolgreich durchführen kann. Du wirst dich mit der Technik von Suchmaschinen und mit verschiedenen Strategien der Informationssuche auseinandersetzen. Neben der Theorie steht die Praxis im Zentrum der Arbeit. Bei jedem Posten kannst du eigene Erfahrungen sammeln und du wirst Aufgaben lösen müssen.

Arbeitsweise

Jeder Posten ist in drei Bereiche aufgeteilt:

Das Symbol  bezeichnet den Theorieteil. Lies ihn aufmerksam durch und streiche wichtige Passagen an. Wenn du etwas nicht verstehst, dann frage die Lehrperson.

Das Symbol  steht für eigene Experimente am Computer. Hier werden dir einige Angaben gegeben mit denen du selbständig am Computer Erfahrungen sammeln kannst. Es ist wichtig, dass du die angeregten Untersuchungen seriös durchführst.

Letztlich steht das Symbol  für Testaufgaben, die du selbständig lösen sollst. Alle Antworten sind schriftlich festzuhalten und anschliessend der Lehrperson zur Korrektur zu zeigen.

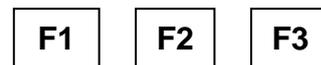
Die ganze Werkstatt ist in 6 Ebenen gegliedert, wobei jede Ebene aus einem oder mehreren Posten besteht. Die Reihenfolge der Posten auf jeder Ebene ist beliebig. Von einer Ebene darf erst auf die nächste übergegangen werden, wenn alle Posten bearbeitet wurden, weil jede nächste Ebene auf dem Wissen der vorhergehenden Ebene(n) aufbaut. Die X-Ebene ist freiwillig. Die U-Ebene beinhaltet ein Übungsblatt mit Fragen die zu lösen sind. Es dient als Praxistest. Die Korrektur der U-Ebene findet bei der Lehrperson statt.

Verwende die folgende Grafik gleichzeitig als Arbeitskontrolle, in dem du die behandelten Posten durchstreichst.

1. **I – Ebene** Internet allgemein



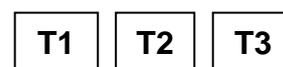
2. **F – Ebene** Funktionsweise von Suchsystemen



3. **S – Ebene** verschiedene Suchdienste



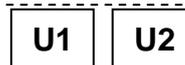
4. **T – Ebene** Technik des Tsuchens



5. **X – Ebene** freiwillige X Posten



6. **U – Ebene** Uebungen



Themen der einzelnen Posten:

- I1** Das Internet – unendliche Weiten
- I2** Informationsdienste
- F1** Die Rangliste
- F2** Der Index
- F3** Die Relevanz
- S1** Interne Suchmaschinen
- S2** Spezielle Suchmaschinen
- S3** allgemeine Suchmaschinen
- S4** Katalogdienste
- S5** Metasuchdienste
- T1** allgemeine Suchtipps
- T2** Die UND-Strategie
- T3** iterative Suchtechnik
- X1** Google – Ein Pionier
- X2** Relevanzrückkoppelung
- U1** Übungen
- U2** Übungen

Zeitbedarf

Zur Bearbeitung stehen dir insgesamt 8-10 Lektionen zur Verfügung. Das ergibt ca. 2 Posten pro Lektion. Dies ist knapp. Es ist deshalb empfehlenswert, vor der Lektion jeweils den Theorieteil der Posten bereits durchzulesen, um sich während den Lektionen auf die praktische Arbeit konzentrieren zu können.

Hinweis

Du darfst die Unterlagen zu den Posten behalten und sie somit auch beschreiben.

Die Datei *Linkliste.html* enthält die notwendigen Links pro Posten!

Wir wünschen dir bei der Bearbeitung der Werkstatt viel Spass!

I1 Das Internet - unendliche Weiten



Die Entstehung des Internets geht auf das Jahr 1969 zurück. Damals begann das amerikanische Verteidigungsmysterium (DoD) das sogenannte ARPA-Net aufzubauen. Das Netz verband verschiedene militärische Stützpunkte (technisch wollen wir sie Knoten nennen) miteinander. Neu an der Struktur war, dass dieses Netz keine Zentrale besass. Jeder Knoten musste dafür schauen, dass er für ihn bestimmte Datenpakete herauspickte und alle übrigen an seine Nachbarknoten weitergab. Das DoD erhoffte sich damit einen besseren Schutz seiner Anlagen bei einem gegnerischen Angriff, denn auch durch den Ausfall einzelner Knoten konnte das Netz sehr wohl noch weiter funktionieren. Zudem wurde die Spionage erschwert, da man nicht exakt wusste, welchen Weg ein Datenpaket durch das Netz der Knoten nahm.

In den folgenden Jahren fand auch die Wissenschaft Interesse an dieser Art Netzwerk und es wurden nach und nach amerikanische Universitäten nach dem Vorbild des ARPA-Net verbunden. Dadurch wurde der Datenaustausch unter den Wissenschaftlern wesentlich vereinfacht.

In den 70er und 80er Jahren stieg die Anzahl Netzwerkteilnehmer und -knoten immer mehr. Trotzdem blieb das Netz in diesen Jahrzehnten hauptsächlich der Wissenschaft vorenthalten. Dies hatte einen einfachen Grund: Es gab damals noch keine Browser¹, mit denen man auf simple Weise von einem PC aus durchs Netz „surfen“ konnte.

Mitte der 90er Jahre startete die rasante Ausbreitung der Nutzung des Internets auch im Heimbereich. Vorgängig waren aber zwei Erfindungen nötig. Zum einen entwickelte das CERN (European Organization for Nuclear Research) in Genf 1989/1990 die Seitenbeschreibungssprache HTML², zum Anderen entstand die erste Browser-Software. Mit HTML wurde es möglich, (Internet-)Seiten zu (be)schreiben, ohne dabei viel Speicherplatz zu benötigen. Diese mit HTML beschriebenen Seiten konnten dann übers Internet bezogen und auf dem eigenen PC dargestellt werden. Die Browser vereinfachten dieses Prozedere indem der Bezug und die Darstellung der Seiten vom Browser automatisch erledigt wird. Dies alles führte zur Nutzung des Internet, wie wir es heute kennen.

Etwas zur Namensgebung: Der Begriff „Web“ stammt aus dem Englischen und heisst eigentlich „Gewebe“ resp. „Spinnennetz“. Der Name ist aus der Struktur des Internets entstanden, die mit ihren unzähligen Knoten ohne Zentrale einem Spinnennetz gleicht. Das Wort „Internet“ setzt sich aus zwei Teilen zusammen, nämlich aus „inter“ (lateinisch für "zwischen") und „net“, der Abkürzung für "networking" (englisch für "vernetzen").

¹ Browser, "Stöberer"

Programm, das benutzt wird, um sich in einem Datensystem oder -netz zu bewegen und zurechtzufinden. Ein Webbrowser ermöglicht den Zugang zu und das Betrachten von grafischen Internet-Seiten (nicht aber deren Bearbeitung!). Bsp. Netscape Navigator, Internet Explorer...

² „Hypertext Markup Language“ etwa: Hypertext-Auszeichnungssprache
Seitenbeschreibungssprache zum Erstellen eines Dokuments im WWW.



Nach diesem kurzen geschichtlichen Überblick sollst du nun selber noch etwas über die Geschichte des Internets nachlesen. Auf den folgenden Internetseiten wird die Internet-Geschichte beschrieben:

<http://www.phil-fak.uni-duesseldorf.de/mmedia/web/start.html> (12.12.2002)

http://userpage.chemie.fu-berlin.de/~sunny/Internet_Arbeit/net_nindex.html (12.12.2002)

<http://www.psychologie.uni-bonn.de/sozial/staff/musch/history.htm> (12.12.2002)

<http://www.michaelkaul.de/Geschichte/zakon/zakon.html> (12.12.2002)

In der Einführung wurde bewusst auf genaue Zahlen verzichtet. Versuche anhand der angegebenen Adressen dir ein Bild zu machen, wie das Internet bis heute gewachsen ist. Suche exakte Zahlen und vergleiche die einzelnen Quellen miteinander.



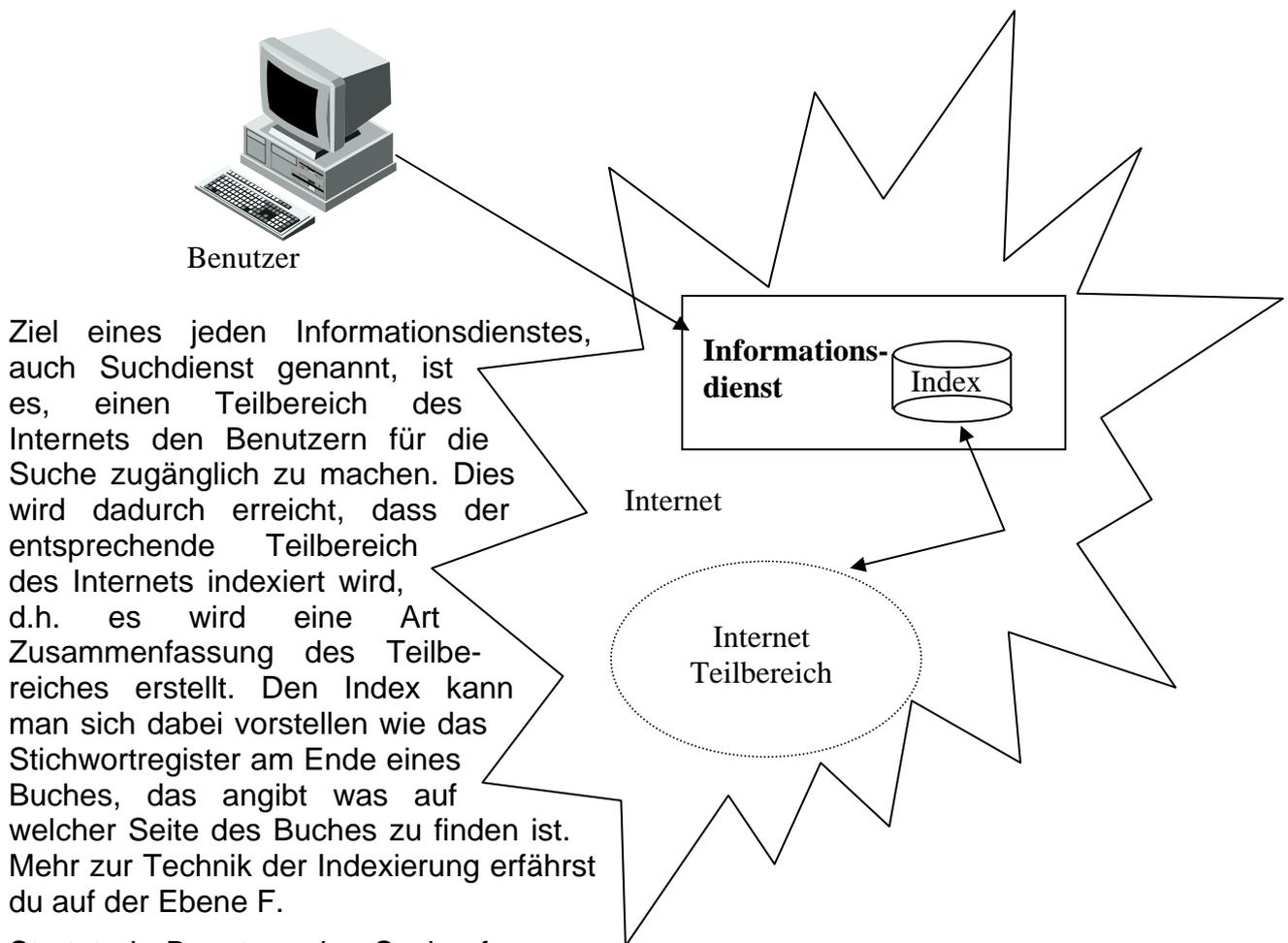
Notiere zu jedem der folgenden Begriffe kurz, ob es wohl sinnvoll ist etwas auf dem Internet zu suchen oder nicht. Begründe jede Antwort kurz. Achtung: Es geht nicht darum, die Antwort auf die gestellte Frage zu finden, sondern nur ob das Internet ein geeignetes Medium zur Beantwortung ist oder nicht!

1. Informationen für einen Geschichtsvortrag über „Das Leben der Römer“
2. Wann fährt der nächste Zug von Sarnen nach Interlaken?
3. Was läuft heute Abend im Kino Capitol 1 in Luzern?
4. Was steht auf der herausgerissenen Seite 65 meines als Occasion gekauften Chemiebuches?
5. Leider habe ich Hans Müller, der mit mir die Matura gemacht hat, aus den Augen verloren. Wo lebt er wohl heute?
6. Für den Französischunterricht brauche ich einen französisch-gesprochenen Nachrichtenbeitrag.

I2 Informationsdienste



Die folgende Grafik zeigt die Charakteristik von Informationsdiensten im Internet



Startet ein Benutzer eine Suchanfrage an das Informationssystem, so schaut das System nach, ob es in „seinem“ Teilbereich des Internets Seiten hat, die zur Suchanfrage passen. Da das Informationssystem aus Zeitgründen nicht jedes Mal den ganzen Teilbereich durchsuchen kann, schaut es nur in seinem Index nach.



Mit der rasanten Zunahme des Angebots im Netz steigt auch die Zahl der Informationsdienste schnell an. So gibt es praktisch für jeden erdenklichen Teilbereich des Internets einen speziellen Suchdienst. Die Seite www.suchlexikon.de (12.12.02), sie ist eine Art Suchdienst für Suchdienste, listet aktuell (am 12.12.02) 2695 deutschsprachige Informationsdienste auf.

Folgenden Internetadressen repräsentieren alle einen Informationsdienst. Versuche herauszufinden, welches Teilgebiet des Internets jeder dieser Suchdienste abdeckt. Schau dir dazu die Beschreibungen bei den einzelnen Suchdiensten an.

<http://www.webcam-center.ch> (12.12.02)

http://www.schule.bayern.de/scripts/basis/m_abf.idc (12.12.02)

www.altavista.de (12.12.02)

<http://www.woman.de/suche/index.html> (12.12.02)

www.google.de (12.12.02)

<http://www.all4travel.de/suche.html> (12.12.02)



Starte die bereits oben zitierte Seite www.suchlexikon.de (12.12.02). Versuche Suchdienste zu finden, die die angegebenen Teilbereiche des Internets abdecken.

1. Konzerte, Musikveranstaltungen, Musikgruppen, Komponisten und Kompositionen etc. Teste den Suchdienst, indem du z.B. Informationen zu deiner Lieblingsmusikgruppe und/oder deren Konzertdaten suchst.
2. Wetterdaten im Internet. Teste den Suchdienst, indem du aktuelle Wetterdaten aus verschiedenen Städten abrufst..

F1 Die Rangliste



Jede Suchanfrage wird vom Suchdienst mit einer Liste der Treffer beantwortet. Diese Liste kann von 0 bis zu mehreren Millionen Treffer beinhalten. In der Folge wollen wir uns ansehen, nach welchen Kriterien die Rangliste zustande kommt, d.h. warum ist ein gefundenes Dokument in der Rangliste weiter vorne oder weiter hinten.

Nach welchen Regeln ein Suchdienst die Rangliste erstellt ist eigentlich Betriebsgeheimnis des jeweiligen Informationsdienstes. Dies aus gutem Grund: Wenn ein Suchdienst die Kriterien und Gewichtungen bekannt gibt, nach welcher er die zutreffenden Dokumente ordnet, so kann das von findigen Programmierern ausgenutzt werden, um ihre Seiten möglichst zuoberst auf der Rangliste erscheinen zu lassen.

Trotzdem gibt es einige Regeln, die von den meisten Suchdiensten mit mehr oder weniger Gewicht berücksichtigt werden.

Prinzip 1: „Je mehr Suchbegriffe im Dokument vorkommen, desto wichtiger ist das Dokument.“

Bsp: Suchanfrage „kantonsschule obwalden“

... Auf das neue Schuljahr wurde die Studentafel der **Kantonsschule Obwalden** geändert. Englisch ist nun bereits ab der 1. Klasse obligatorisch. ...

... Bezüglich der Vernehmlassung zur Armee XXI äusserte sich die Regierung des Kantons **Obwalden** wie folgt: ...

Das linke Dokument erscheint in der Rangliste weiter oben, da beide Suchbegriffe im Dokument enthalten sind.

Prinzip 2: „Je häufiger ein Suchbegriff im Dokument vorkommt, umso wichtiger ist das Dokument.“

Bsp: Suchanfrage „erdbeben“

... In China ereignete sich ein **Erdbeben** mit Epizentrum 200 km nördlich der Hauptstadt. Das **Erdbeben** erreichte die Stärke 7.3 auf der Richterskala. Die durch das **Erdbeben** verursachten Schäden gehen in die Millionen...

... Nach dem **Erdbeben** im Jahre 1990 wurden die Schäden zügig repariert. Die Stadt erlebte seither einen wirtschaftlichen Aufschwung...

Das linke Dokument erscheint in der Rangliste weiter oben, da der Suchbegriff im Dokument häufiger vorkommt. Achtung: Zuviel ist ungesund! Die meisten Suchdienste ignorieren Seiten, auf denen ein Begriff zu häufig vorkommt.

Prinzip 3: „Dokumente mit seltenen Suchbegriffen sind wichtiger als Dokumente mit häufig vorkommenden Begriffen“

Bsp: Suchanfrage „flugzeug unglück lockerbie“

... **Flugzeug** bei **Lockerbie** abgestürzt. Die Behörden sprechen von einer Katastrophe...

..**Flugzeug** in Kambodscha abgestürzt. Das **Unglück** forderte 63 Todesopfer...

In beiden Dokumenten kommen gleichviele Suchbegriffe vor. Da der Begriff „Lockerbie“ aber viel seltener ist als der Begriff „Unglück“, erscheint das linke Dokument weiter oben in der Rangliste.

Prinzip 4: „Ein kurzes Dokument ist wichtiger als ein langes Dokument.“

Bsp: Suchanfrage „nelson mandela“

Friedensnobelträger von 1993

Der Friedensnobelpreis 1993 ging an Nelson Mandela, geboren am 25.7.1918 in Transkei, Südafrika. Er trat 1944 dem ANC bei und engagierte sich in der Anti-Apartheid-Politik.

Friedensnobelpreisträger

1990 ...
1991 ...
1992 ...
1993 Nelson Mandela, F.W. de Klerk
1994 ...

Das linke, wesentlich kürzere Dokument erscheint in der Rangliste weiter oben als die lange Liste aller Nobelpreisträger.

Prinzip 5: „Je näher die Suchbegriffe beieinander liegen, umso wichtiger ist ein Dokument.“

Bsp: Suchanfrage „kantonsschule obwalden“

... Auf das neue Schuljahr wurde die Studentafel der **Kantonsschule Obwalden** geändert. Englisch ist nun bereits ab der 1. Klasse obligatorisch. ...

Uri: Franz Zraggen, **Kantonsschule Uri**; ...
Obwalden: Urs Imfeld, Stiftsschule Engelberg;
...

Im linken Dokument liegen die Suchbegriffe direkt nebeneinander, es erscheint deshalb weiter oben in der Rangliste.

Prinzip 6: „Je weiter vorne ein Suchbegriff im Dokument auftaucht, desto wichtiger ist das Dokument“

Bsp: Suchanfrage „pythagoras“

Pythagoras von Samos (ca. 540 - 500 v. u. Z.):
Pythagoras ist ein Schüler von Themistokleia und vielleicht auch des Pherekydes. Er war Mann der Theano von Kroton sowie Vater von Arignote, Myia, Damo, Telauges und Mnesarchos...

Lehrplan Mathematik
geometrische Objekte...
das Dreieck...
Satz des Pythagoras...

Im linken Dokument steht der Suchbegriff ganz am Anfang des Dokumentes, deshalb erscheint es in der Rangliste weiter oben als das rechte Dokument.



Analog zu den Rangierungsprinzipien 1 – 6 sind noch weitere ähnliche Regeln denkbar. Formuliere 2 zusätzliche Prinzipien und diskutiere sie mit einem Mitschüler/einer Mitschülerin, der/die sich auch gerade bei der Frage befindet. Im Zweifelsfall fragt die Lehrperson.



Wähle den Suchdienst www.fireball.de (12.12.02) und gib als Suchanfrage „uboot kursk untergang“ an. Schau dir die ersten 5 Treffer an. Welche der Prinzipien zur Erstellung der Rangliste kannst du eindeutig erkennen? Erkenne mindestens 3 verschiedene.

F2 Der Index



Im Posten I2 hast du gelernt, dass jeder Informationsdienst einen Teil des Internets für die Suche zugänglich macht. Wird dem Dienst eine Suchanfrage gestellt, so schaut der Suchdienst in seinem Index nach, ob er ein (oder mehrere) dazu passendes Dokument verzeichnet hat.

An diesem Posten wollen wir uns ansehen, wie sich ein Suchdienst seinen Index aufbaut, d.h. wie er sich merkt, was im für ihn wichtigen Teilbereich des Internets für Web-Seiten vorhanden sind.

Zuerst muss der Suchdienst möglichst viele Seiten aufstöbern, die zum untersuchten Teilbereich im Internet gehören. Dies geschieht dadurch, dass der Informationsdienst kleine Roboter (eigentlich sind es kleine Programme) ins Internet aussendet. Die Roboter, auch Bots genannt, „sehen“ sich die Seiten im Internet an. Wie ein Surfer gehen sie von Link zu Link und somit von einer Seite zur anderen. Bei jeder Seite, die der Suchdienst in seinem Index haben möchte, wird das folgende Verfahren angewendet.

Beispiel: Wir betrachten die folgende (fiktive) Internet-Seite www.mondlandung.ch, die dem Index des Suchdienstes hinzugefügt werden soll:

Am 20. Juli 1969 landeten die Amerikaner erstmals auf dem Mond. Neil Armstrong berührte den Mond als Erster und verkündete: „Ein kleiner Schritt für einen Menschen, aber ein Riesenschritt für die Menschheit.“

Zuerst wird festgestellt, um welche Sprache es sich handelt. Da die Seite in Deutsch ist, wird nun eine **Buchstabenumwandlung** gemacht. ä, ö, ü wird zu ae, oe, ue.

Am 20. Juli 1969 landeten die Amerikaner erstmals auf dem Mond. Neil Armstrong beruehrte den Mond als Erster und verkuendete: „Ein kleiner Schritt fuer einen Menschen, aber ein Riesenschritt für die Menschheit.“

Anschliessend wird die **Wortextraktion** durchgeführt, d.h. es werden die Satzzeichen weggelassen.

Am 20 Juli 1969 landeten die Amerikaner erstmals auf dem Mond Neil Armstrong beruehrte den Mond als Erster und verkuendete Ein kleiner Schritt fuer einen Menschen aber ein Riesenschritt für die Menschheit

Nun werden alle **Stoppwörter eliminiert**. Stoppwörter sind Begriffe, die nichts oder nur sehr wenig zur Beschreibung des Inhalts beitragen. Bsp. am, die, auf,...

20 Juli 1969 landeten Amerikaner erstmals Mond Neil Armstrong beruehrte Mond Erster verkuendete kleiner Schritt Menschen Riesenschritt Menschheit

In der nächsten Phase werden die **Wörter zerlegt und normalisiert**. So wird etwa aus „Riesenschritt“ „riesig“ und „Schritt“. Bei der Normalisierung wird lediglich der Wortstamm, ohne Endungen und verschiedene Schreibweisen, übernommen. Z.B. statt „beruehrte“ „beruehr“. Zudem werden alle Gross- in Kleinbuchstaben umgewandelt.

20 juli 1969 land amerika erst mal mond neil armstrong beruehr mond erst verkuend klein schritt mensch riesig schritt menschheit

Die so zurechtgemachte Seite wird nun analysiert und dem Index beigefügt. Dazu werden die einzelnen Begriffe im Dokument gezählt und in einer Tabelle aufgelistet.

Begriff	Häufigkeit	Positionen
mond	2	8, 12		
erst	2	6, 13		
schritt	2	16, 19		
1969	1	3		
...		

Der Suchdienst weiss nun, welche Begriffe sich auf der Seite www.mondlandung.ch befinden. Wird nun eine Suchanfrage wie beispielsweise „mondlandung erste“ gestellt, so erkennt der Suchdienst, nachdem er die Suchanfrage ebenfalls in „mond“ „landung“ „erst“ zerlegt hat, dass die Seite www.mondlandung.ch dieser Anfrage entspricht und liefert sie als Treffer zurück. Wo in der Rangliste die Seite dann erscheint, hängt von den Rangierungsprinzipien ab. Vgl. Posten F1.

Das gezeigte Verfahren ist nur ein Muster für die Erstellung eines Indexes. Nicht alle Suchdienste wenden alle Teile an, oder sie verwenden noch andere Mechanismen. Im Experimentierteil weiter unten wirst du bei einigen Suchdiensten versuchen herauszufinden, ob der Index nach dem gezeigten Muster erstellt wird.



Überlege dir, wie du bei einem Suchdienst herausfinden kannst, ob er Wortzerlegung, Wortnormalisierung, Stoppworteliminierung und Umlauterkennung durchführt.

Untersuche dann bei den beiden Suchdienste www.google.de (12.12.02) und www.fireball.de (12.12.02) ob sie...

... Wortzerlegung durchführen. Suchanfrage „toskanareise“

... Wortnormalisierung durchführen. Suchanfrage „landen“

... Stoppwörter eliminieren. Suchanfrage „sein und haben“ resp. „sein haben“

... Umlaute/Accent erkennen. Verwende als Suchanfrage „nestlé“ resp. „nestle“



Das Schweizerische Parlament (www.parlament.ch 26.04.06) verwendet für seine Seiten einen Suchdienst mit dem Namen RotondoSpider. Untersuche diesen auf Wortzerlegung, Wortnormalisierung, Stoppworteliminierung und Umlauterkennung. Dokumentiere deine Suchanfragen und erkläre deine Interpretation.

F3 Die Relevanz

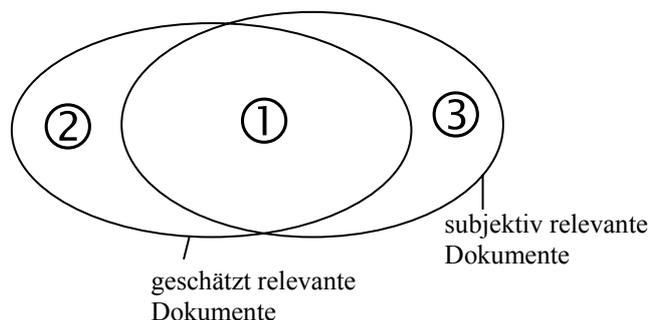


Gemäss Langenscheidts Fremdwörterbuch bedeutet *Relevanz* soviel wie *Wichtigkeit, Bedeutsamkeit* und *relevant* meint *erheblich, bedeutsam, wichtig, wesentlich*. Im Zusammenhang mit Suchdiensten unterscheiden wir zwei Arten von Relevanz die sogenannte **subjektive Relevanz** und die **geschätzte Relevanz**.

Stellt ein Benutzer an ein Suchsystem eine Anfrage, so hat er gewisse Erwartungen, was die Resultate betrifft. Auf Grund dieser Erwartungen kann er entscheiden, ob ein Trefferdokument für ihn relevant ist oder nicht. Diese vom Benutzer abhängige Wichtigkeit nennen wir subjektive Relevanz.

Der Suchdienst dagegen muss auf Grund der gestellten Anfrage entscheiden, welche Dokumente in seinem repräsentierten Teilbereich des Internets relevant sein könnten. Der Informationsdienst muss also die Relevanz von Dokumenten bezüglich der Anfrage abschätzen. Dies eben ist die geschätzte Relevanz. Für diese Relevanzabschätzung verwendet der Suchdienst unter anderem die Rangierungsprinzipien wie sie im Posten F1 erklärt sind.

Im Idealfall decken sich die subjektive und die geschätzte Relevanz. D.h. der Suchdienst liefert genau jene Seiten als Treffer, die für den Benutzer wichtig sind. In der Praxis ist das aber nie der Fall. Die Menge aller für den Benutzer relevanten Dokumente (subjektive Relevanz) und die Menge aller Dokumente die der Suchdienst für wichtig erachtet (geschätzte Relevanz) haben zwar eine gemeinsame Schnittmenge (Bereich ①), aber es gibt immer Seiten, die als Treffer vom Dienst geliefert werden, die aber für den Benutzer nicht brauchbar sind (Bereich ②). Andererseits werden vom Suchdienst gewisse Seiten nicht als relevant erachtet, und somit nicht als Treffer angegeben, die aber für den Benutzer sehr wohl wichtig wären (Bereich ③).



Durch die Wahl des richtigen Suchdienstes und geschickt gestellte Suchanfragen kann der Benutzer selber sehr viel dazu beitragen, dass die Schnittmenge möglichst gross ist. Dabei ist es wichtig, dass der Benutzer versteht, wie Suchdienste arbeiten. Mehr über Suchdienste erfährst du in der Ebene S dieser Werkstatt. Wie man geschickte Anfragen stellt, ist Gegenstand der Ebene T.



Ca. 75% der Internetbenutzer haben keine Ahnung, wie man Suchdienste effizient verwendet, d.h. ihre Suchanfragen sind derart unpräzise gestellt, dass die Schnittmenge der subjektiv relevanten Dokumente und der geschätzt-relevanten Dokumente nur klein ist. Dies ist zwar eine ziemlich gewagte Behauptung, aber am

Ende dieses Experimentes wirst du erkennen, dass es vermutlich ziemlich genau zutrifft.

Beim Suchdienst Fireball lässt sich beobachten, wie andere Benutzer suchen. Wenn du www.fireball.de (12.12.02) aufrufst und anschliessend auf *Live-Suche* klickst, bekommst du jede Minute die neuste Liste der soeben an Fireball gestellten Anfragen.

Schau dir die Suchanfragen an. Wenn du auf eine der aufgelisteten Suchanfragen klickst, werden die zu dieser Suchanfrage gehörenden Treffer aufgelistet.

Wähle einige Suchanfragen aus. Überlege dir was der Benutzer wohl suchen wollte, d.h. was für ihn subjektiv relevant sein könnte. Schau dann die Liste der dazu gelieferten Treffer an. Versuche abzuschätzen ob die Suchanfrage gut war, oder ob sich die Menge der subjektiv-relevanten und die Menge der geschätzt-relevanten Dokumente nur schwach schneiden. Mit anderen Worten, bekommt der Benutzer was er wohl wollte, oder ist das Resultat unbrauchbar.



Ziehe die Konsequenzen aus den oben gemachten Überlegungen. Formuliere schriftlich ca. drei Kriterien, die eine schlechte Suchanfrage ausmachen. Gemeint ist, wenn eine Suchanfrage von der beschriebenen Art ist, so werden verhältnismässig nur wenig subjektiv-relevante Dokumente als Treffer angezeigt.

Erkläre bei jedem Kriterium warum das eine schlechte Anfrage ist resp. warum kaum eine gute Trefferliste herauskommt.

S1 Interne Suchmaschinen



Mittels internen Suchmaschinen lassen sich alle zu einer Domain gehörenden Web-Seiten durchsuchen. So haben zum Beispiel Hompages von Zeitungen, Firmen, Instituten etc. meistens irgendwo einen Knopf oder eine Eingabezeile mit dem Titel „Suche“. Bei der Startseite der Schweizerischen Eidgenossenschaft befindet sich rechts oben gut sichtbar eine Eingabezeile für Suchanfragen. So können zum Beispiel die Seiten der Bundesverwaltung auf das Stichwort „Mutterschaftsversicherung“ durchsucht werden. Das interne Suchsystem liefert als Treffer alle Seiten, die der Suchanfrage entsprechen und sich auf den Servern der Eidgenossenschaft befinden.

Diese Suche lässt sich vergleichen mit dem Befehl „Suchen“ im Menü Extras des Windowsexplorers. Mit ihm lassen sich die lokalen Festplatten nach Dateien durchsuchen. Mit internen Suchmaschinen analog dazu die „lokalen Seiten“, die zu einer Domain gehören.

Es ist wichtig interne Suchmaschinen von allgemeinen Suchmaschinen (vgl. Posten S3 resp. S4) zu unterscheiden. Mit internen Suchmaschinen lässt sich nicht das ganze Internet durchsuchen. So macht es z.B. wenig Sinn, in der internen Suchmaschine der Bundesseiten den Begriff „mp3“ einzugeben, weil man gerne auf dem Internet Musikdateien herunterladen möchte. Der Bund bietet auf seinen Seiten wohl kaum mp3-Musikdateien an.

Interne Suchmaschinen sind meist mit den modernsten Suchtechnologien ausgestattet, die es erlauben die Seiten in möglichst geringer Zeit effizient zu durchsuchen. Solche Technologien sind etwa Wortnormalisierung beim Erstellen des Indexes und beim Zerlegen der Anfrage (vgl. Posten F2) oder Relevanzrückkoppelung (vgl. Posten X2). Zudem sind interne Suchmaschinen frei von störenden Werbebannern.

Der Grund für diese beiden Begebenheiten ist einfach. Wenn der Hersteller einer Homepage sich entschliesst auf seinen Seiten eine interne Suchmaschine einzusetzen, so programmiert er sie in den wenigsten Fällen selber. Er kauft vielmehr eine Software ein, die ihm diese Suchmöglichkeiten anbietet. Genau wie jemand der Texte mit dem Computer schreiben will sich eine Textverarbeitungssoftware kauft, kauft jemand der auf seinen Seiten eine Suche



durchführen lassen will eine interne Suchmaschine. Deshalb haben die Hersteller solcher Suchmaschinensoftware allen Grund, möglichst die beste (und das heisst hier die schnellste) Maschine anzubieten. Damit verdienen sie ihr Geld. Aus diesem Grund müssen die internen Suchmaschinen auch keine Werbung platzieren.

Ein weiterer grosser Vorteil von internen Suchmaschinen ist, dass sie oft auch Archive durchsuchen. Solche Archive sind meistens nicht eigentliche Webseiten, sondern die Daten aus dem Archiv werden bei Bedarf auf einer dynamisch erstellten Webseite dargestellt. Allgemeine Suchmaschinen, die das ganze Internet durchsuchen, können solche Seiten nicht finden und zu den Archiven haben sie keinen Zugang. Ein typisches Beispiel sind Zeitungsarchive, die mittels einer internen Suchmaschine durchforstet werden können. Leider sind solche Archive immer häufiger kostenpflichtig.



Versuche bei den folgenden Webseiten die interne Suchmaschine zu finden. Starte jeweils einige Suchanfragen um zu sehen, wie die Maschine die Resultate angibt.

www.facts.ch (12.12.02)

www.stadt Luzern.ch (12.12.02)

www.ow.ch (12.12.02)

www.nasa.gov (12.12.02)

www.maxon.ch (12.12.02)

www.swiss.ch (12.12.02)



1. Versuche mit Hilfe der internen Suchmaschine auf der Homepage der Gemeinde Sarnen www.sarnen.ch (12.12.02) herauszufinden, wie der Präsident der Einwohnergemeinde heisst, und wie seine Email-Adresse lautet.
2. Zur Abwechslung mal eine Frage, die sich auf eine englische Seite bezieht: Wähle die interne Suchmaschine für NASA-Seiten www.nasa.gov (12.12.02) und versuche herauszufinden, was das spezielle an der NASA-Mission STS 51-L war.

S2 Spezielle Suchmaschinen



Die Blinde-Kuh Suchmaschine

Jedes Wort Mindestens ein Wort

Wenn du etwas im Internet suchen möchtest, **gib das Wort oder mehrere Wörter ein** und klicke dann mit der Maus auf den Suchen-Button. Wenn du schon ein großer Suchmeister bist, dann klick doch mal oben rechts bei **"Experten"**. Wie man hier so richtig ordentlich suchen kann, erfährst du im [Suchmaschinenkurs](#).

PAPERBALL

ALLE RUBRIKEN POLITIK WIRTSCHAFT SPORT KULTUR LOKALES VERMISCHTES

Suchbegriff eingeben

in folgenden Zeitungen

alle
Aachener Nachrichten
Aachener-Zeitung
Aargauer Zeitung
Aller Zeitung

in den Rubriken

alle
Politik
Wirtschaft
Sport
Kultur

Spezielle Suchmaschinen repräsentieren einen klar definierten, meistens thematisch orientierten, Teilbereich des Internet. In der Fachsprache wird das eine vertikale Datenkollektion genannt.

Im Gegensatz zu allgemeinen Suchmaschinen (vgl. Posten S3) wissen spezielle Suchmaschinen über bestimmte Themen sehr genau Bescheid, da sie sich darauf spezialisiert haben. Es gibt beispielsweise spezielle Suchmaschinen für Kinder (www.blinde-kuh.de 12.12.02), Suchmaschinen für Filme (us.imdb.com/search 12.12.02),

Das Suchlexikon

Das Suchlexikon ist ein durchsuchbares Verzeichnis spezieller deutschsprachiger Suchmaschinen, Kataloge, Verzeichnisse und Linksammlungen.

Suche nach spezieller Suchmaschine: [weitere Suchoptionen](#)

Suchmaschinen für aktuelle Zeitungsartikel (www.paperball.de 12.12.02) oder auch Suchmaschinen für Suchmaschinen, wie etwa www.suchlexikon.de (12.12.02).

Der Unterschied von allgemeinen Suchmaschinen und speziellen Suchmaschinen ist manchmal gar nicht so gross. Die Schweizer Version der Suchmaschine Google (www.google.ch 12.12.02) kann als spezielle Suchmaschine für Schweizer Seiten (*.ch), als spezielle Suchmaschine für deutschsprachige Seiten oder als allgemeine Suchmaschine verwendet werden. Es ist dazu lediglich ein Mausklick nötig.

Google [Erweiterte Suche](#) [Einstellungen](#) [Sprach-Tools](#) [Suchtipps](#)

Suche: Das Web Seiten auf Deutsch Seiten aus der Schweiz

Weitere wichtige Vertreter von speziellen Suchmaschinen sind

jene, die nicht den (textlichen) Inhalt von Internetseiten durchsuchen, sondern die Bilder, Musikdateien oder Videos aufstöbern. Zwei Beispiele sind die Suchmaschine Altavista (www.altavista.com 12.12.02) und Google (www.google.com 12.12.02). Beide bieten die Möglichkeit, spezielle Suchanfragen für Bilder, Musikdateien, Videos etc. zu starten.

altavista [Web](#) [Image](#) [MP3/Audio](#) [Video](#) [News](#)



Spezielle Suchmaschinen erleichtern oft das tägliche Leben, da sie exakt für Bereiche geschaffen sind, in denen man manchmal schnell im Internet etwas nachsehen möchte. Beispiele für solche Bereiche sind:

- Fahrplan des öffentlichen Verkehrs
- Das Wetter in einer Region
- Flugverbindungen
- Online-Einkaufsmöglichkeiten
- Telefonnummern
- Übernachtungsmöglichkeiten

Versuche für die oben aufgelisteten Bereiche spezielle Suchmaschinen zu finden. Als Startpunkt kann dir www.suchlexikon.de (12.12.02) dienen. Notiere dir zu jedem Bereich eine (oder mehrere) Internetadressen von Suchdiensten die dir gegebenenfalls weiterhelfen.



Der norwegische Maler Edvard Munch (1863-1944) malte im Jahre 1908 ein Bild mit dem Namen „der Schrei“ es wurde sein wohl berühmtestes Werk.

Verwende die Bildersuche in www.altavista.com (12.12.02). Hole das Bild „der Schrei“ auf den Bildschirm und zeige es dem Lehrer.

S3 Allgemeine Suchmaschinen



Im Gegensatz zu speziellen Suchmaschinen (vgl. Posten S2) versuchen allgemeine Suchmaschinen möglichst das „ganze“ Internet abzudecken. Sie sind darauf aus, möglichst viele Seiten im Internet zu untersuchen und im Index aufzunehmen. Man sagt, dass allgemeine Suchmaschinen eine horizontale Datenkollektion besitzen. Damit ist gemeint, dass sie nicht primär alles zu einem Thema aufnehmen, sondern eben möglichst einen grossen Teil vom Internet abdecken.

Wie im Posten F2 kennengelernt schicken die allgemeinen Suchmaschinen Roboter aus, die das Internet durchsuchen, um an die (neuen) Seiten heranzukommen, die noch nicht im eigenen Index gespeichert sind. Diese Roboter „lesen“ den ganzen Text auf einer Seite, d.h. sie untersuchen ihn gemäss den Mechanismen der Indexierung. Aus diesem Grund nennt man allgemeine Suchmaschinen auch Volltextsuchmaschinen, weil sie eben den vollen Text von Webseiten untersuchen und damit ihren Index erweitern (vgl. Posten F2).

Allgemeine Suchmaschinen sind meist mit unzähligen Werbebannern bestückt. Der Grund dazu ist einfach. Allgemeine Suchmaschinen verdienen ihr Geld durch die auf ihrer Seite dargestellte Werbung. Die Suchanfrage ist für den Benutzer kostenlos, darum müssen die Betreiber von solchen Suchdiensten ihre Fläche für Werbung zur Verfügung stellen. Diese Abhängigkeit von Werbung führt nicht nur zu störender, oft penetranter Werbung auf den Seiten der Suchmaschinen, sondern verhindern oft auch, dass allgemeine Suchmaschinen mit modernen Suchtechniken ausgestattet werden. Warum? Weil es den Betreibern nicht wichtig ist, die Suchanfrage des Benutzers möglichst rasch zu beantworten. Im Gegenteil: Je länger ein Benutzer sich auf der Suchoberfläche befindet, umso mehr Werbung muss er über sich ergehen lassen.

In einzelnen Fällen geht die Abhängigkeit der Suchmaschinenbetreiber sogar noch weiter. Sie verkaufen nicht nur ihre Werbefläche, sondern für gutes Geld kann eine Firma auch noch bestimmen, welche Seite bei einer Suchanfrage als erstes in der Rangliste erscheinen soll. So wird z. B. eine Fluggesellschaft der Suchmaschine dafür Geld zahlen, dass bei Suchbegriffen wie „Flugzeug“ oder „Fluggesellschaft“ ihre Homepage als erstes in der Rangliste erscheint.



Die Oberflächen von allgemeinen Suchmaschinen sehen alle mehr oder weniger gleich aus. Neben einer Eingabezeile für die Suchanfrage gibt es meistens noch Knöpfe, die zwischen einer einfachen und einer komplexeren Sucheingabe (z.B. Profisuche) wählen lassen. In der Rubrik Hilfe kann der Benutzer nachschauen, wie genau die Suchbegriffe eingegeben werden müssen, d.h. wie die Syntax aussieht. Mehr zu Suchanfragen und wie sie aussehen müssen lernst du auf der T-Ebene.

Allgemeine Suchmaschinen eignen sich besonders für konkrete Suchanfragen, d.h. der Benutzer will etwas ganz bestimmtes in Erfahrung bringen und sich nicht bloss einen Überblick verschaffen.



Experiment 1

Wie eingangs erwähnt ist es das A und O einer allgemeinen Suchmaschine, möglichst viele Internetseiten zu kennen, und somit dem Benutzer möglichst viele relevante Treffer anzugeben. Wir wollen die Datenmengen einiger Suchdienste vergleichen. Wähle die folgenden Dienste und gib jeweils den Suchbegriff „schule“ ein. Schreibe bei jeder Suchmaschine auf, wie viele Webseiten sie gefunden hat.

www.abacho.de (12.12.02)

www.altavista.de (12.12.02)

www.excite.de (12.12.02)

www.fireball.de (12.12.02)

www.kolibri.de (12.12.02)

www.lycos.de (12.12.02)

www.nathan.de (12.12.02)

Experiment 2

Verwende die deutsche Version der Suchmaschine Altavista (www.altavista.de 12.12.02). Gib einzelne Suchbegriffe aus verschiedenen Bereichen ein. Z.B. „golf“, „flug“ etc. Vergleiche nun die Werbung am rechten Rand der Rangliste und die ersten Resultate in der Rubrik der Sponsoren. Du wirst feststellen, dass sie wenn immer möglich der Suchanfrage angepasst werden.



Vor einigen Jahren wurde der damals amtierende Schachweltmeister von einem Computer geschlagen. Versuche mit Hilfe der Suchdienste von Experiment 1 folgende Fragen zu beantworten:

1. Wie hiess der Schachcomputer?
2. Wie hiess der Weltmeister?
3. Wann und wo fand die Partie statt?
4. Wie lautete das Schlussresultat?
5. Wo und wann spielte letztmals ein amtierender Schachweltmeister gegen einen Computer? Wie endete die Begegnung?

S4 Katalogdienste



Der Index von allgemeinen Suchdiensten lässt sich mit einem Stichwortverzeichnis vergleichen (vgl. F2 und S3). Im Index ist aufgelistet, welcher Suchbegriff sich wie häufig auf welcher Webseite befindet.

Der Index von Katalogdiensten gleicht dagegen eher einem Branchentelefonbuch, in dem zu jeder Branche die entsprechenden Firmen aufgelistet sind. In Katalogdiensten werden die Webseiten thematisch in Rubriken (sog. Kategorien oder Verzeichnisse) eingeteilt. Diese Einteilung erfolgt „von Hand“, d.h. durch Redaktoren, die die Seiten begutachten und dann den Rubriken zuordnen, oder durch den Ersteller der Webseite selber, der dem Katalog angibt, welcher (oder welchen) Rubrik(en) seine Seite zuzuordnen ist.

Die Rubriken sind hierarchisch geordnet. Jede Rubrik hat wieder einige Unterrubriken. Das nebenstehende Bild zeigt die Rubriken des bekanntesten Katalogdienstes (Yahoo).

Katalogdienste werden typischerweise für Suchanfragen verwendet, bei denen es darum geht sich einen Überblick zu verschaffen, z.B. Infos über Städte, spezielle Musikbereiche etc.

Bildung & Ausbildung Hochschulen , Schulen , Ausbildung ...	Nachrichten & Medien Top Themen , TV , Zeitschriften , Zeitungen ...
Computer & Internet Internet , Hardware , Software , Chat ...	Nachschlagewerke Bibliotheken , Telefonnummern , Lexika ...
Geisteswissenschaften Psychologie , Geschichte , Sprachen ...	Naturwiss. & Technik Tiere , Biologie , Handy und SMS ...
Gesellschaft & Soziales Umwelt , Religion , Sexualität , Gay ...	Sport & Freizeit Autos , F1 , Fußball , Spiele , Reisen ...
Gesundheit Medizin , Krankheiten , Psyche ...	Staat & Politik Politik , Behörden , Recht , Steuern ...
Handel & Wirtschaft Firmen , Finanzen , Jobs , Immobilien ...	Städte & Länder Dt. Städte , Länder , Italien , Spanien ...
Kunst & Kultur Kunst , Literatur , Theater ...	Unterhaltung Coole Links , Musik , Humor , Kino ...

Es gibt zwei Möglichkeiten, wie bei einem Katalogdienst relevante Seiten gefunden werden können. Erstens kann der Benutzer sich von Rubrik zu Rubrik durchklicken, bis er bei den Seiten angelangt ist, die ihn interessieren. Oder er kann die Eingabezeile für Suchanfragen benutzen. Der Katalog liefert ihm dann ganze Rubriken, die der Suchanfrage entsprechen, und zusätzlich noch einzelne Seiten. Der Benutzer kann auch die Suchanfrage in der Eingabezeile auf eine bestimmte Rubrik

Gefundene Kategorien (1 - 1 von 1) Städte und Länder > Länder > Schweiz > Kantone • Obwalden
Gefundene Web-Sites (1 - 13 von 13) Städte und Länder > Länder > Schweiz > Kantone > Obwalden > Regionalführer • Kanton Obwalden - Kurzportrait, Parlament, Verwaltung, Gerichte, Gemeinden und Wirtschaft. • Obwalden Net - Informationen über den Kanton - Gemeinden - Firmen - Vereine und Veranstaltungen

beschränken.

Volltextsuchmaschinen und Katalogdienste wachsen immer mehr zusammen. So haben heutzutage die meisten allgemeinen Suchmaschinen auch eine Verzeichnisstruktur (z.B. Google). Der Benutzer kann wählen, welche Art der Suche er anwenden möchte.



Schaue dir die folgenden deutschsprachigen Kataloge näher an. Vergleiche insbesondere die Verzeichnisstruktur. Gibt es markante Unterschiede in den Top-Rubriken? Vergleiche auch die vorhandene Datenmenge, indem du in allen Katalogen nach dem gleichen Suchbegriff (z.B. „volleyball“) suchst.

www.yahoo.de (12.12.02)

dir.web.de 12.12.02)

www.dino-online.de (12.12.02)

www.allesklar.de (12.12.02)



Du möchtest am kommenden Wochenende nach München verreisen. Dich interessieren folgende Informationen:

1. Wie ist das Wetter aktuell in München.
2. Welche Verbindungen zur Hin- und Rückreise stehen dir offen?
3. Wie kannst du möglichst günstig übernachten?
4. Was kostet ein 24h-Ticket für die öffentlichen Verkehrsmittel?
5. Was geht den am Samstagabend so ab?

Wähle einen Katalogdienst und versuche damit diese Fragen zu beantworten.

S5 Metasuchdienste



Im online-Internetwörterbuch steht unter „meta search engine“ (dt. Metasuchmaschine resp. Metasuchdienst) folgende Definition:

„Übergeordnete Suchmaschine, die gleichzeitig mehrere Datenbanken bzw. Einzel-Suchmaschinen nach den eingegebenen Suchbegriffen durchforstet. Damit erspart man sich oft die aufwändige Suche nach der benötigten Information über die einzelnen Suchmaschinen.“

Eine Metasuchmaschine hat somit keinen eigenen Index. Sie schickt die gestellte Suchanfrage an mehrere allgemeine Suchmaschinen und fasst deren Resultate zusammen. Dadurch wird natürlich die untersuchte Datenkollektion sehr umfangreich, da sich die einzelnen Suchdienste ergänzen.

Bei der Oberfläche unterscheiden sich

MetaGer, die MetaSuchmaschine
über deutschsprachige Suchmaschinen

Ein Service des [RRZN&RVS, Universität Hannover](#)

Geben Sie einfach ein oder mehrere Suchwörter ein:

Alle Wörter sollen im Dokument vorkommen

Metasuchdienste kaum von allgemeinen Suchmaschinen. Auch sie besitzen eine Eingabezeile für die Suchanfrage. Bei den meisten Metasuchdiensten kann angeklickt werden, an welche Suchmaschinen der Metasuchdienst seine Anfragen schicken soll.

Die von uns ausgewählten besten Suchdienste: (Falls Sie Voreinstellungen unter diesen ausgewählten Diensten ändern wollen, klicken Sie bitte den entsprechenden Schalter an.)				
<input checked="" type="checkbox"/> Netfind	<input checked="" type="checkbox"/> Altavista	<input checked="" type="checkbox"/> Yahoo	<input checked="" type="checkbox"/> web.de	<input checked="" type="checkbox"/> AltaVista.de
<input checked="" type="checkbox"/> Lycos	<input type="checkbox"/> Crawler.de	<input checked="" type="checkbox"/> Witch	<input checked="" type="checkbox"/> T-Online	<input type="checkbox"/> Walhello
<input type="checkbox"/> MSN	<input type="checkbox"/> Speedfind	<input type="checkbox"/> Dino	<input type="checkbox"/> AllesKlar	<input type="checkbox"/> Intersearch.at

Bei einigen Metasuchdiensten wird bei den Resultaten angegeben, welche allgemeine(n) Suchmaschine(n) den Treffer geliefert hat. Als zusätzliches

Rangierungsprinzip bei Metasuchmaschinen kommt dazu, dass je mehr der angefragten Suchmaschinen eine Seite gefunden haben, umso relevanter wird sie für den Benutzer sein. Die gezeigten Bildschirm Ausschnitte wurden vom deutschsprachigen Metasuchdienst *Metager* (meta.rrzn.uni-hannover.de 12.12.02) gemacht. Als Suchanfrage diente „kantonsschule obwalden“.

1) *** Lehrer & Lehrerinnen Verein Obwalden
QCheck: http://www.lvo.ch/stufen/berichte.htm
<ul style="list-style-type: none">• (gefunden von: Lycos) Lehrerinnen und Lehrerverein Stufen Letzte Aktualisierung am Sa., 02.06.2001 Bei Stufenverantwortl. und AVS 09.11.2000 Kindergarten• (gefunden von: AltaVista) Lehrerinnen und Lehrerverein Stufen. Letzte Aktualisierung am Don, 08.02.2001. F• (gefunden von: AltaVista.de) Lehrerinnen und Lehrerverein Stufen. Letzte Aktualisierung am Don, 08.02

Die Anwendung von Metasuchmaschinen empfiehlt sich vor allem dann, wenn man mit allgemeinen Suchmaschinen und Katalogdiensten kein befriedigendes Resultat erhalten hat. Ebenfalls sinnvoll ist die Verwendung von Metasuchdiensten, wenn zu einem sehr beschränkten Themenkreis, zu dem es nur wenige Informationen auf dem Netz gibt, möglichst alles gefunden werden soll.



An dieser Stelle sollst du dich mit einigen Metasuchdiensten vertraut machen. Gib bei den folgenden Diensten als Suchanfrage deinen Vornamen und deinen Nachnamen ein. Damit siehst du, ob im Internet irgendwas von dir, oder von jemandem mit deinem Namen, vorhanden ist. Solltest du nichts Interessantes finden, wo versuche es mit dem Namen deiner Lehrperson.

Wähle folgende deutschsprachigen und internationalen Metasuchdienste:

www.metacrawler.com (12.12.02)

www.highway61.com (12.12.02)

www.dogpile.com (12.12.02)

www.apollo7.de (12.12.02)

meta.rrzn.uni-hannover.de (12.12.02)

www.megaspider.com (12.12.02)



Versuche mit Hilfe von Metasuchdiensten die folgenden Fragen zu beantworten:

1. Wer ist Martin Glinz und wo arbeitet er? Wie sieht er aus?
2. Wer behauptet, dass Frau Nuss in Bümpliz wohnt? Mit welchem Buchstaben beginnt der Vorname von Frau Nuss?
3. Welchen Jahrgang hat der Schauspieler Daniel Radcliffe? Welche Rollen spielte er?

T1 allgemeine Suchtipps



Tipp 1: Wähle den richtigen Suchdienst

Wie du in der S-Ebene gelernt hast, ist eine allgemeine Suchmaschine nicht das gleiche wie ein Katalogdienst, interne Suchmaschinen decken ganz andere Bereiche ab als spezielle Suchmaschinen und Metasuchdienste bieten noch einmal eine andere Voraussetzung. Es lohnt sich gut zu überlegen, welchen Dienst man in Anspruch nehmen will.

Tipp 2: Die Sprache beachten

Eine Suchanfrage in Deutsch, liefert keine Seiten in anderen Sprachen. Deshalb ist es oft nützlich eine Anfrage in verschiedenen Sprachen zu stellen, sofern man auch in der Lage ist eine Seite in der entsprechenden Sprache zu verstehen. In Suchdiensten eingebaute Übersetzungsprogramme können für die weitere Suche nützlich sein, aber nicht um den Inhalt einer Seite für die weitere Verwertung zu übersetzen. Beachte auch, dass eine vermeintlich Deutsch gestellte Anfrage durchaus auch Englisch sein kann und somit wohl die meisten Treffer Englische Seiten sein werden (z.B. „internet homepage swisscom“)

Tipp 3: Rechtschreibung beachten

Bei Umlauten und Sonderzeichen müssen auch alternative Schreibweisen probiert werden. Denke auch daran, dass das Suchsystem sich unter einem Wort nichts vorstellen kann. Es weiss beispielsweise bei der Suchanfrage „band“ nicht, ob nach Musikgruppen, mehrteiligen Bücherreihen oder Strumpfbändern gesucht werden soll.

Tipp 4: Sag dem Suchdienst was du weisst, nicht was du suchst

Suchsysteme können keine Gedanken lesen. Sie können nur schauen, ob in der Suchanfrage enthaltene Begriffe auf Internetseiten vorkommen. Je mehr Informationen der Benutzer dem Suchsystem zukommen lässt, desto treffendere Antworten kann das System liefern.

Tipp 5: Achte auf die Syntax der Suchdienste

Nicht alle Suchdienste verlangen eine gleich-formatierte Eingabe. Meistens bedeutet die Eingabe von mehreren Wörtern eine UND-Verknüpfung, – aber nicht immer. Die wichtigsten Eingaberegeln betreffend der UND-Verknüpfung, der ODER-Verknüpfung, der Phrase (d.h. mehrere Wörter werden zusammen als eines angeschaut z.B. „Alex Birrer“) und der Trunkierung (d.h. wird ein Suchbegriff auch als Teil eines Wortes gefunden oder nicht) können im Online-Help der jeweiligen Suchmaschine nachgeschaut werden.

Tipp 6: Suche mit Strategie und /oder Erfahrung

Auf der T-Ebene dieser Werkstatt lernst du verschiedene Suchtechniken kennen. Zielgerichtetes Suchen braucht neben etwas technischem Wissen

hauptsächlich Erfahrung. Versuche deshalb diese Verfahren möglichst anzuwenden um damit deine eigene Sucherfahrung zu erlangen.

Tipp 7: Interpretiere die Rangliste

Die 2-3 ersten Treffer sollten unbedingt genau angesehen werden. Das erste Dokument kann je nach Suchdienst mit Werbung zu tun haben, aber spätestens ab dem 2. Dokument müsste es in Richtung relevanter Seiten gehen. Sollte das nicht der Fall sein, so ist sofort die Suchanfrage anzupassen oder eine andere Suchmaschine zu wählen. Schau dir Dokumente, die wenig bis gar nichts versprechen, überhaupt nicht an. Du verlierst nur Zeit damit.

Tipp 8: Gefunden heisst noch nicht brauchbar

Nicht jede Seite, die subjektiv-relevant erscheint, hat zwingend auch einen korrekten Inhalt. Bevor Internetquellen weiter verarbeitet werden, ist unbedingt die Qualität der Seite sicherzustellen. Jedermann kann bekanntlich jeglichen Nonsens im Internet publik machen.

Tipp 9: Internetadresse erraten

Oft ist es geschickt, bevor du eine Suchanfrage an einen Suchdienst stellst zu versuchen die Internetadresse zu erraten. Dieses Verfahren ist insbesondere bei Firmen, Verwaltungen etc. lohnenswert.



Versuche folgende Internetadressen zu erraten:

Kanton Obwalden	Stadt Luzern	Zürich Versicherungen
Weisses Haus	Schw. Eidgenossenschaft	Musikgruppe U2
Schweizer Fernsehen	Internationales Rotes Kreuz	Offizielle Vatikan-Homepage



1. Endungen von Internetadressen

Im Verlaufe dieser Werkstatt hast du bereits viele Internetseiten angeschaut. Es wird Zeit, uns mal etwas näher mit Internetadressen zu befassen. Besonders interessieren sollen uns an dieser Stelle die Endungen, in der Fachsprache Top-Level-Domains genannt. Das sind die letzten Buchstaben vor dem Ende der Adresse resp. vor dem ersten Schrägstrich (www.ksobwalden.ch oder www.nasa.gov).

Versuche mit Hilfe des internets herauszufinden, was die folgenden Endungen zu bedeuten haben: *com, ch, edu, gov, org, net, fr, es, tm* und *ie*.

2. Eingabesyntax

Vergleiche bei den folgenden Suchdiensten www.altavista.de (12.12.02), www.google.de (12.12.02), www.fireball.de (12.12.02) und www.allesklar.de (12.12.02) wie UND-, ODER-Verknüpfungen, Phrasen und Trunkierung eingegeben werden müssen. Schau in den Helps der einzelnen Suchdienste nach. Erstelle eine Tabelle mit den Resultaten.

T2 Die UND-Strategie



Die UND-Strategie ist ein Vertreter der bool'schen Suchverfahren, benannt nach George Boole (1815-1864). UND ist ein sogenannter bool'scher Operator. Weitere solche Operatoren sind ODER und NICHT. Der Bool'sche UND-Operator lässt sich umgangssprachlich leicht verstehen: „Aussage1 UND Aussage2“ heisst, dass das Ganze wahr ist, wenn beide Aussagen wahr sind. Bsp. „Wir gehen auf die Schulreise wenn es nicht regnet und wenn die SBB nicht streiken“ gemeint ist, dass beides erfüllt sein muss, dass die Schulreise stattfindet.

Genauso sind Internetrecherchen mit der UND-Strategie zu verstehen. Es werden bei der Suchanfrage mehrere Suchbegriffe angegeben und mit UND verknüpft. Dies veranlasst den Suchdienst, Seiten zu suchen, auf denen alle Suchbegriffe vorkommen, also Suchbegriff1 UND Suchbegriff2 UND Suchbegriff3 ... Fehlt nur ein einzelner Suchbegriff auf der Seite, so wird diese Seite in der Rangliste der Treffer nicht aufgelistet. Die meisten Suchdienste verwenden automatisch die UND-Verknüpfung, wenn mehrere Wörter als Suchanfrage eingegeben werden.

In der folgenden Grafik enthalten die Seiten 1 und 2 alle eingegebenen Begriffe, somit werden diese beiden Seiten als Treffer zurückgegeben, die Seite 3 aber nicht, es fehlt der Begriff *Informatik*.

Suchanfrage*:

Seite 1	Seite 2	Seite 3
... studierte an der ETH Zürich und unterrichtet zur Zeit Informatik an der Kantonsschule Obwalden in Sarnen....	... im Bildungskonzept des Kantons Obwaldendass das Fach Informatik auch in den Schulen der Stufe Sek II (Kantonsschule , Berufsschule)...	... Chor der Kantonsschule Obwalden unter der Leitung.....begeisterte das Publikum mit ihren modernen...

*Die hier verwendete Syntax (d.h. Schreibweise) für die UND-Verknüpfung entspricht derjenigen der meisten Suchmaschinen, wie z.B. Altavista oder Google. Vgl. Posten T1 und die Hilfe-Seiten der entsprechenden Suchmaschinen.

Die Anzahl Treffer bei einer Suche mit UND-Verknüpfung nimmt meistens mit zunehmender Anzahl Suchbegriffe ab. Die Treffermenge ist die Schnittmenge aller Resultatmengen der einzelnen Suchbegriffe.

Wird zu einer bestehenden Suchanfrage noch ein weiteres Wort hinzugeführt, so bedeutet das automatisch, dass von den vorher angegebenen Suchresultaten nun nur noch jene angezeigt werden, die auch das neue Wort beinhalten.

Die Verknüpfung mit UND ist immer dann sinnvoll, wenn der Benutzer genau weiss, welches Wort resp. welche Worte auf den gesuchten Seiten vorkommen sollen.

Anstelle einer UND-Verknüpfung kann es ausnahmsweise auch von Vorteil sein, Suchbegriffe mit einem ODER zu verknüpfen. Die Suchmaschine liefert dann alle Seiten als Resultat, die mindestens eines der angegebenen Worte enthalten. Die Treffermenge ist damit natürlich bedeutend grösser, da sie die Vereinigungsmenge

aller Resultatmengen der einzelnen Begriffe ist. Die Verknüpfung von Suchbegriffen mit ODER ist dann sinnvoll, wenn der Benutzer nicht genau weiss welches Wort auf der Seite steht, oder wenn zu einem Begriff viele Synonyme existieren.



Verwende die drei Suchdienste www.fireball.de (12.12.02), www.google.de (12.12.02) und www.allesklar.de (12.12.02). Suche mit Hilfe von UND-Verknüpfungen nach Seiten zu deiner Liebblingsschauspielerin oder deinem Liebblingsschauspieler oder zu deiner Liebblingsserie/-sendung.



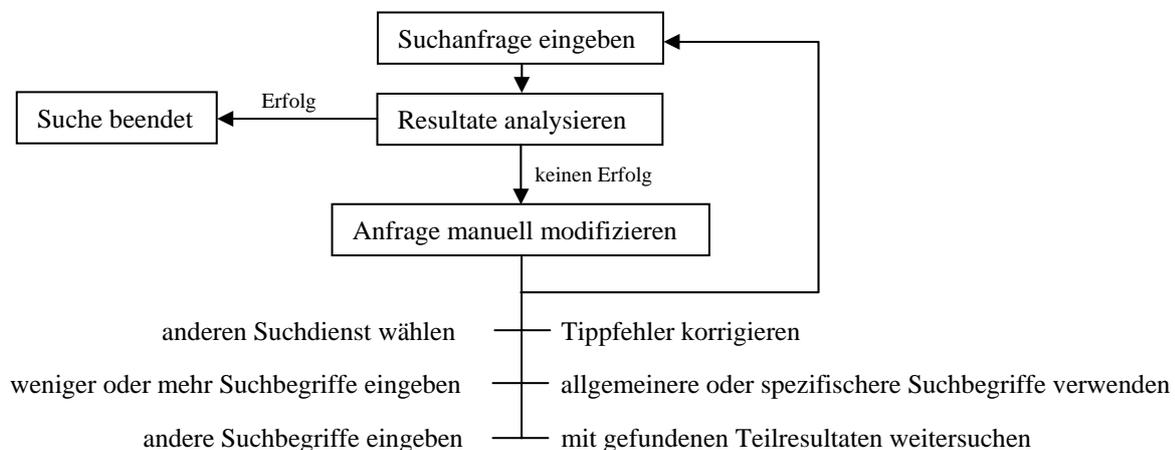
1. Wie heisst der flächenmässig grösste (künstlich angelegte) Stausee der Schweiz? Welche Fläche hat er?
2. Wie heisst der hellste Stern im Sternbild des Skorpions? Woher hat er seinen Namen?
3. In der Verfilmung eines Romans von Grisham spielt der Newcomer Matt Damon einen jungen Anwalt. Wie heisst der Film und wer war der Regisseur (Name, Vorname und Geburtsjahr).
4. Bei den Olympischen Spielen 2000 in Sidney hat der australische Schwimmer Ian „Thorpedo“ Thrope drei Goldmedaillen gewonnen und wurde so zum Liebling der Nation. Einige Jahre früher hat ein Schwimmer an Olympia noch wesentlich mehr Goldmedaillen errungen. Wann und wo war das? Wie hiess der Schwimmer und wie viele goldene Auszeichnungen gewann er?
5. Im Gegensatz zur Vorderseite von Euromünzen, die bei allen Länder der Währungsunion gleich ist, ist die Gestaltung der Rückseite der Münzen jeder Nation überlassen. Ein Land wählte als Rückseite eine berühmte Proportionsstudie aus der Antike. Auf welcher Münze befindet sich die Darstellung (Wert und Land)?
6. Beschreibe das Gemälde von Edouard Manet, das die Flucht von Henri Rochefort zeigt. Wann wurde das Gemälde gemalt? Wo steht es heute?

T3 Iterative Suchtechnik



Informationssuche ist ein iterativer, das heißt in mehreren Durchgängen auszuführender Prozess. Diese Aussage gilt in den meisten Fällen, denn nur selten führt die erste Suchanfrage an einen Suchdienst gleich zum Ziel. Im Normalfall arbeitet man sich schrittweise zum gewünschten Resultat vor. Mit jeder Anfrage lernt man etwas hinzu und verwendet das zusätzliche Wissen für die folgenden Suchanfragen.

Die folgende Grafik zeigt schematisch, wie eine iterative Suche aussehen könnte:



Die häufigste Korrektur der Anfrage besteht darin, dass mehr, weniger oder andere Suchbegriffe genommen werden müssen. Damit kann die Treffermenge beeinflusst werden.

Wir betrachten das folgende Beispiel:

Wir möchten herausfinden, was am kommenden Samstag Abend in Sarnen los ist. Zur Suche wählen wir die UND-Strategie und die Suchmaschine www.google.ch (12.12.02). Die erste Eingabe lautet „veranstaltung sarnen samstag abend“. Die Liste der Treffer (98 an der Zahl) ist nicht sehr vielversprechend. Sie enthält vorab Seiten, die irgend welche Veranstaltungen beschreiben oder dokumentieren, die an einem Samstag Abend in Sarnen stattgefunden haben. Auf jeden Fall nichts Interessantes zum kommenden Samstag Abend.

Fazit: Die Suchanfrage mit der UND-Strategie wahr wohl falsch gestellt. Insbesondere die Begriffe „samstag abend“ scheinen in die falsche Richtung zu gehen. Was ist zu tun? Wir verkleinern die Anzahl Suchbegriffe und geben nur noch „veranstaltung sarnen“ ein, in der Hoffnung damit auf einen Veranstaltungskalender von Sarnen zu stossen, in dem wir dann selber nachschauen können, was am kommenden Samstag los ist.

Die Suchanfrage liefert zwar 1370 Treffer, aber bereits das erste Dokument hat den Titel „Sarnen Tourismus – Veranstaltungen“. Die genauere Betrachtung der Seite zeigt, dass sie genau unserem Bedürfnis entspricht.

X1 Google – Ein Pionier



Die Suchmaschine Google ist in kürzester Zeit vom Geheimtipp zur am meisten verwendeten allgemeinen Suchmaschine im Internet aufgestiegen. Deshalb verdient Google ein eigenes Kapitel. Google gibt es in den Varianten www.google.com, www.google.de und auch www.google.ch.

In diesem Posten wird anhand von verschiedenen Originalartikeln der Werdegang von Google aufgezeigt.

- | |
|--|
| Artikel 1: „ Warum Google? “
Selbstwerbung des Suchdienstes |
| Artikel 2: „ Suchmaschine Google ist der Retter genervter Surfer “
Berliner Morgenpost, 3.3.2001 |
| Artikel 3: „ Die Erfolgsstory von Google “
World online |

Artikel 1:
Selbstwerbung des Suchdienstes

Warum Google?

Weil Google schnell die relevantesten Suchergebnisse liefert!

Mit Hilfe der ausgefeiltesten wegweisenden Technologien findet Google schnell die richtigen Ergebnisse für jede Suchanfrage. Google gibt relevante Ergebnisse zurück, weil die gefundenen Web-Sites automatisch anhand der Link-Struktur des Internets bewertet werden.

Der überwältigende Umfang der Informationen im Web erfordert einen hervorragenden Suchdienst, der diese Informationen zugänglich macht und den Benutzern erschließt. Ohne ein leistungsfähiges Suchwerkzeug kann die Suche nach einer bestimmten Web-Site so schwierig sein wie die Suche nach einem Buch in einer Bibliothek ohne Katalog, in der die Bücher ohne Ordnungsprinzip willkürlich in die Regale gestellt werden.

Google bringt Ordnung ins Web

Google ordnet das Informationschaos. Google ist ein Suchdienst, wie er sein soll: kein redigiertes, eingeschränktes Verzeichnis oder eine Liste von Ergebnissen, die an den Meistbietenden versteigert wurden, sondern eine durchdachte Methode, das Internet anhand seiner eigenen Struktur zu organisieren.

Google revolutioniert die Web-Suche mit seiner zum Patent angemeldeten PageRankTM-Technologie. PageRank nutzt die Struktur des Webs, die durch die Art definiert wird, wie eine Web-Seite direkt, unmittelbar und ohne Drittparteien mit anderen Web-Seiten verknüpft werden kann. In einer gewissen Weise demokratisiert diese Link-Struktur das Internet auf automatischem Wege. Sie enthierarchisiert die Informationen und ermöglicht den ungehinderten Fluss der Informationen und Ideen zwischen den einzelnen Sites.

So funktioniert PageRank

PageRank nutzt die umfassende Link-Struktur des Webs als

Organisationskriterium und profitiert so von dessen einzigartiger demokratischer Beschaffenheit. Im wesentlichen interpretiert Google einen Link von Seite A zu Seite B als eine Stimme *von* Seite A *für* Seite B. Google beurteilt die Wichtigkeit einer Seite nach der Anzahl der abgegebenen Stimmen. Google berücksichtigt jedoch nicht nur die Anzahl der Stimmen bzw. Links, sondern analysiert auch die Seite, von der die Stimme ausgeht. Stimmen von Seiten, die selbst als "wichtig" eingestuft werden, haben eine größere Bedeutung bei der Bewertung der Wichtigkeit anderer Seiten.

Diese wichtigen, qualitativ hochwertigen Ergebnisse erhalten einen höheren PageRank-Wert. Dieser bestimmt, an welcher Stelle in der Liste der Ergebnisse die zurückgegebene Seite erscheint. Auf diese Weise ist der PageRank-Wert ein Indikator für die allgemeine Wichtigkeit einer Seite, der nicht von einer bestimmten Suchanfrage abhängt, sondern eine Eigenschaft der Seite selbst darstellt. Er wird aufgrund von Daten aus dem Web bestimmt, die Google mittels komplexer Algorithmen, die die Link-Struktur bewerten, analysiert.

Natürlich ist eine wichtige Seite für Sie wertlos, wenn Sie nicht Ihrer Suchanfrage entspricht. Daher wendet Google ausgeklügelte Textanalysetechniken an, um sowohl wichtige als auch für Ihre Suchanfrage relevante Seiten zu finden. Wenn Google eine Seite analysiert, wird beispielsweise auch berücksichtigt, welche Informationen zu dieser Seite auf den Seiten enthalten sind, die auf die analysierte Seite verweisen.

Integrität

Die komplexen, automatisierten Suchtechniken von Google schließen jeden menschlichen Eingriff aus. Im Gegensatz zu anderen Suchmaschinen verhindert die Struktur von Google, dass jemand einen höheren PageRank-Wert erwerben kann oder die Ergebnisse auf andere Weise kommerziell verfälscht. Eine Google

Suche ist eine ehrliche und objektive Methode, auf einfache Weise qualitativ hochwertige Web-Sites zu finden.

Google gibt nur die Seiten zurück, welche die eingegebenen Suchbegriffe enthalten.

Im Gegensatz zu vielen anderen Suchmaschinen gibt Google nur solche Ergebnisse zurück, die entweder im Seitentext oder in den Namen der Links, die auf die Seite verweisen, alle Suchbegriffe enthalten. Sie brauchen sich nie wieder mit Ergebnissen herumzürgeren, die gar nichts mit den von Ihnen eingegebenen Suchbegriffen zu tun haben.

Google beachtet die Position Ihrer Suchbegriffe auf einer Seite.

Die Ergebnisse von Google enthalten nicht nur alle eingegebenen Suchbegriffe; Google analysiert auch, wie nahe diese Begriffe auf einer Seite beieinander stehen. Im Gegensatz zu vielen anderen Suchmaschinen bewertet Google die Relevanz der Ergebnisse auch anhand der Nähe der Suchbegriffe. Wir bewerten Ergebnisse, bei denen die Suchergebnisse nahe beieinander liegen, höher, damit Sie sich nicht durch eine Unmenge von irrelevanten Ergebnissen arbeiten müssen.

Google bietet eine aussagekräftige Vorschau für jedes Ergebnis.

Statt feststehender Zusammenfassungen zeigt Google in den Suchergebnissen einen Auszug aus dem Text an, der auf Ihre Suchanfrage abgestimmt ist, und in dem Ihre Suchbegriffe hervorgehoben sind. Dies erspart Ihnen die Zeit und den Ärger, irrelevante Web-Seiten zu laden.

Mit Google haben Sie Glück!

Google ist unübertroffen darin, das richtige Ergebnis für eine allgemeine Suchanfrage (wenn der Suchbegriff zum Beispiel ein Firmenname ist) an erster Stelle zurückzugeben. Wir vertrauen dieser Funktion so sehr, dass wir die Schaltfläche

"Auf gut Glück™!" eingerichtet haben, die Sie direkt zur Web-Site des ersten Suchergebnisses führt. Mit der Funktion "Auf gut Glück!" von Google gelangen Sie so schnell wie möglich an die gewünschten Informationen.

Google speichert Web-Seiten im Cache.

Google speichert in seinem Cache zahlreiche Kopien von Web-Seiten, auf die Sie bei einem zeitweiligen Ausfall des Servers der Web-Seite zurückgreifen können. Wenn der Server nicht verfügbar ist, kann die Kopie der benötigten Seite im Cache von Google Ihr Retter sein. Häufig können Sie auf das Material im Cache schneller zugreifen als über den Link zur eigentlichen Seite; die auf diese Weise erhaltenen Informationen sind jedoch möglicherweise nicht auf dem neuesten Stand. In den meisten Fällen ist das aber immer noch besser als ein "404 Not Found"-Fehler!

Artikel 2:

Berliner Morgenpost, 3.3.2001

Suchmaschine Google ist der Retter genervter Surfer

Von Christiane Schulzki-Haddouti

Eine kleine, unscheinbare Suchmaschine hat sich in wenigen Monaten vom Geheimtipp zur größten und intelligentesten Suchmaschine im Netz entwickelt: Google. Dabei liest sich die Geschichte der Firma wie die berühmte Amerika-Story vom Aufstieg des Tellerwäschers zum Millionär.

Auf den ersten Blick sieht Google wie ein Suchdienst irgendeiner Universität aus. Bis auf ein einsames Eingabefeld und das Logo ist nichts zu sehen. Keine Kataloge, keine Nachrichten, kein Übersetzungsdienst, keine Werbung - kein Portal. Google kennt nur eins: das Suchen von Informationen. Und, erstaunlich für den Erstnutzer, oft stehen die relevanten Ergebnisse an erster Stelle. Der Trick: Google bewertet die Webseiten als wichtig, auf die viele andere Webseiten

verweisen. Die am häufigsten «Verlinkten» stehen ganz oben.

Entwickelt wurde Google vom Computer-Ingenieur Lawrence Page und dem Mathematiker Sergey Brin. Sie lernten sich an der Stanford University kennen, wo sie «einen Prototypen für eine umfassende Suchmaschine» entwickelten. Heute ist Google ein florierendes Unternehmen. Etwa 40 der 200 Mitarbeiter haben einen Dokortitel, die Hälfte ist im Software-Engineering tätig. Ohne Werbung, nur durch Mund- oder besser Netzpropaganda entwickelte sich Google rasch zum Geheimtipp. Im Januar 2001 lag die Firma mit 8,8 Millionen Besuchern bereits auf Platz 25 der US-Websiteliste von Media-Matrix und rangiert unter den beliebtesten Suchmaschinen auf Rang fünf. [...].

Vor einem Jahr war der Google-Index noch 150 Millionen Seiten lang. Im Winter 2000 holte er auch im Indizierungswettbewerb gegen Konkurrenten wie Altavista und Inktomi auf. Jetzt erfasst sie die meisten Websites überhaupt: Im November wurden 602 Millionen Webseiten von ihr laut Searchenginewatch.com indiziert. Altavista kannte hingegen gerade einmal 350 Millionen. Da Google jedoch auch Links berücksichtigt, erfasst es auch Seiten, die es selber nie besucht hat. Damit deckt es rund 1,3 Milliarden Seiten ab - im Juli 2000 wurde der Umfang des Webs auf zwei Milliarden Seiten geschätzt. [...].

Das Google-Archiv belegt komprimiert heute mehrere hundert Terabyte Festplattenplatz. Der älteste Eintrag im Index ist rund 30 Tage alt. Praktisch: Neben dem Index speichert Google auch die Webseiten selbst, um bei den Treffern die entsprechenden Suchbegriffe in einer Art Kurzzusammenfassung darstellen zu können.

Hinter Googles schlichter Fassade steckt ein gewaltiger Hardware-Park: Rund 7000 Computer sind auf fünf Standorte an der West- und Ostküste der USA verteilt. Sie bilden die Hardware-Ausstattung von Google, laufen unter dem freien

Betriebssystem Linux und werden von einer kleinen Firma extra für Google assimiliert. Hardware und Software sind für die Parallelverarbeitung ausgelegt. 70 Millionen Abfragen muss die Suchmaschine inzwischen jeden Tag verkraften. Anders als die Portale Altavista.com oder Yahoo.com fungiert der Suchdienst selbst als riesiges Werbeportal für die Firma Google - und sonst für niemanden. Google lizenziert seinen Suchdienst an andere Firmen, wie zum Beispiel Cisco, die ihre eigenen Websites damit erschließen. Aber auch Internetportale haben sich bereits zu einer Google-Lizenz durchgerungen. Im letzten Sommer entschloss sich sogar das amerikanische Yahoo!-Portal, seine Suchmaschine mit Google-Technik auszurüsten - und sich von Inktomi zu verabschieden. Für Farzad Nazem, Chef der Entwicklungsabteilung bei Yahoo!, war klar: Google bietet derzeit die beste Suchtechnologie. Für manche Nutzer reduziert sich das Suchen derzeit auf zwei Klicks: Bei Google suchen, bei Altavista-Babelfish übersetzen.

Das Geheimnis des Erfolgs ist, dass die Google-Entwickler Brin und Page nicht auf Schlüsselwörter (Altavista) oder persönliches Ranking (Yahoo!) setzen, sondern auf Links. Sie schufen Karten, die die Vernetzung über Hyperlinks wiedergaben. Auf diese Weise entsteht ein Page-Ranking, vergleichbar dem Bewertungssystem unter Wissenschaftlern: Je häufiger seine Artikel und Werke zitiert werden, umso angesehener wird er. Und: Die Zitate sind auch ein Hinweis auf die Qualität dieser Werke. Gleiches gilt im Netz. Das Page-Ranking von Google widerspiegelt aber auch das Webverhalten: Surfen ist oft zufälliges Springen von einer zur anderen Seiten auf der Flucht vor Langeweile. Kombiniert mit anderen Methoden macht das die Stärke von Google aus. So wird im Ranking auch berücksichtigt, welche Seiten zu einer anderen Seite verweisen. Dabei bestimmt die Ranghöhe dieser Seiten wiederum die Ranghöhe der anderen Seiten. Wenn eine

Seite zum Beispiel nur einen einzigen Link von Yahoo! erhält, liegt sie bei den Google-Suchergebnissen weiter vorn. Und: Die Suche erfasst nicht nur Text-Links, sondern auch Verweise auf Bilder und Programme. Google verfügt noch über weitere Features. Es merkt sich die Hits und generiert darüber Suchwahrscheinlichkeiten. Zudem gewichtet es auch visuelle Darstellungen: Wörter, die fett oder größer dargestellt sind, werden höher gewichtet als andere Wörter. Seit kurzem kann Google auch alle PDF-Dokumente durchsuchen und ist mit 13 Millionen PDF-Dateien sogar eine bessere Alternative als Adobes eigener Suchservice mit nur einer Million Dateien. Der Clou: Die PDF-Dateien sind bei Google als Text-Dateien abrufbar.

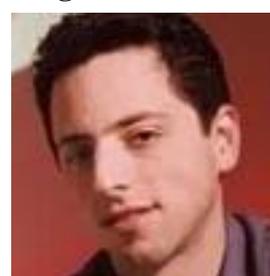
Artikel 3: World online

Die Erfolgsstory von Google

Die Suchmaschine Google wurde im Jahre 1998 an der Stanford Universität von zwei jungen Informatikern entwickelt und hatte Ende September 1999 ihren ersten Online-Auftritt.



Sergey Brin und Larry Page, so die Namen der beiden Google-Schöpfer, ist es gelungen, eine raffinierte Technologie zu entwickeln, mit der auf jede Suchanfrage so rasch und genau wie möglich eine Antwort gefunden wird.



Google sortiert die Suchergebnisse, indem es die Wichtigkeit jeder Webpage mit mathematischen Methoden bewertet. Das bedeutet einen Check von über 500 Millionen Variablen und 2 Milliarden Begriffen (!).

Mathematik und Genauigkeit

Die Besonderheit dieser neuen Technologie, die „Page Rank“ genannt wird, besteht in der Bewertung der Anzahl und der Qualität der Links, die sich auf eine Seite beziehen. Nachdem die inhaltliche Suche erfolgreich abgeschlossen ist, werden die Ergebnisse nach der Wichtigkeit der Webseiten geordnet, die auf die gefundenen Sites verweisen.

So findet sich z.B. eine Wirtschaftsanalyse eines vollkommen Unbekannten auf die von einer anderen unbekanntem Quelle verwiesen wird, am Ende der Ergebnisliste, während sich an der Spitze Artikel von bekannten Wirtschaftsgurus befinden, auf die direkt von der New York Times verwiesen wird. Google betrachtet die Links, die eine Webseite mit einer anderen verbinden als demokratisches Urteil, das von der verweisenden Seite über jene Seite getroffen wird, auf die verwiesen wird. Aus der Gesamtheit aller dieser „Bewertungen“ im Internet wird dann schliesslich die Wichtigkeit einer Internetseite berechnet. Auch die jeweiligen Seiten, auf denen sich ein Link befindet, werden auf dieselbe Art und Weise unter die Lupe genommen.

Es versteht sich also von selbst, dass nicht alle Seiten von gleicher Wichtigkeit sind: Das Geheimnis liegt in der Kontrolle der Ursprungsseiten und in der Zuordnung eines mehr oder weniger hohen „Page Rank“ je nach der Wichtigkeit der Webseite, die die Links enthält. Zudem bietet Google seinen Benutzern Suchergebnisse, die genau der Suchanfrage entsprechen. Anders als bei vielen anderen Suchmaschinen werden nämlich nur Webseiten angezeigt, die alle Schlüsselbegriffe enthalten, wobei überdies noch die Nähe der einzelnen Begriffe innerhalb des Dokuments analysiert wird. Die Gesamtheit dieser Überprüfungskriterien ermöglicht es Google, höchstpräzise Antworten zu liefern.

„I’m feeling lucky“

Daneben bietet Google noch eine ungewöhnliche Option, die es den Surfern erlaubt, sich Google blind anzuvertrauen und direkt auf das erste Link der Ergebnisliste zu springen. Diese aussergewöhnliche Option, die es den Usern erlaubt, so schnell wie möglich zu nützlichen Informationen zu kommen, wurde „I’m feeling lucky“ getauft und hat ihren grössten Nutzen bei häufig gestellten Anfragen, wie z.B. wenn es darum geht, die Homepage grosser Unternehmen oder berühmte Produkte zu finden.

29 Hundertstelsekunden

Die hohe Bearbeitungsgeschwindigkeit ist ein weiteres Markenzeichen von Google. Mit einer Durchschnittsdauer von 29 Hundertstelsekunden pro Anfrage zählt die Suchmaschine zu den schnellsten im ganzen Internet. Ende Juni konnte Google überdies die Erfassung von über einer Milliarde Webseiten vermelden (davon 560 Millionen vollständig) und damit Inktomi als Yahoo-Lieferant ablösen, was die Suche von Sites betrifft, die sich nicht in den Yahoo-Archiven befinden.

Klar und übersichtlich

Die Besonderheit, die an der Homepage von Google am meisten ins Auge sticht, ist ihre Einfachheit. Wer bis dato an die Unübersichtlichkeit anderer Suchmaschinen gewöhnt war, wird zweifelsohne positiv überrascht von der Klarheit und der spartanischen Eleganz von Google sein. Angesichts des Erfolgs von Google haben nun auch andere Suchmotoren damit begonnen, ihre Homepages nach dem minimalistischen Vorbild Googles zu gestalten.

X2 Relevanzrückkoppelung

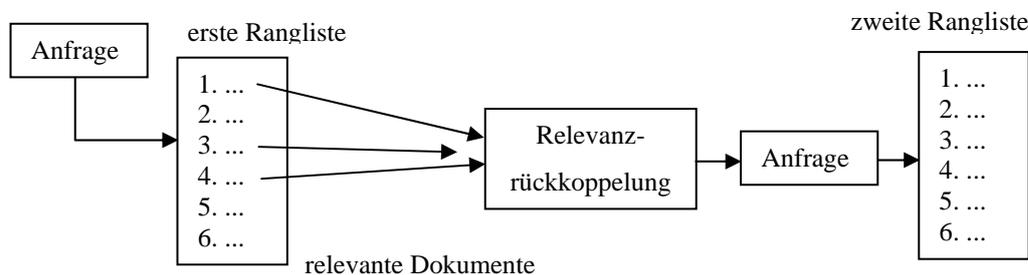


Die Relevanzrückkoppelung stellt ein mächtigeres Werkzeug dar, als die Suche nach ähnlichen Dokumenten (vgl. Posten T3): Nach einer ersten Anfrage beurteilt der Benutzer zunächst, welche der Dokumente in der Rangliste für sein Informatikbedürfnis relevant sind. Diese Information teilt er anschliessend dem Suchsystem mit.

Das Suchsystem untersucht die relevanten Dokumente, identifiziert häufige Suchbegriffe und stellt daraus eine neue Anfrage zusammen. Bestimmte Suchbegriffe werden also als besonders relevant erkannt. Genauso kann das System andere Begriffe als besonders irrelevant erkennen. Schliesslich wird die neue Anfrage verarbeitet und dem Benutzer eine neue Rangliste präsentiert.

Mit Hilfe der Relevanzrückkoppelung lassen sich umfangreiche Anfragen zusammenstellen, mit denen häufig viele zusätzlich relevante Dokumente gefunden werden. Dokumente werden auch dann gefunden, wenn sie keinen einzigen Suchbegriff der ursprünglichen Anfrage enthalten.

Schematisch sieht das wie folgt aus:



Die interne Suchmaschine Eurospider ist einer von wenigen Diensten, die die Relevanzrückkoppelung bereits eingebaut haben. Die Site www.parlament.ch (12.12.02) verwendet den Rotondospider als interne Suchmaschine. Bei jedem vom Suchsystem angegebenen Treffer hat es ein Kästchen zum Anklicken. Damit können die relevanten Dokumente bezeichnet werden und mit „neue Suche“ wird automatisch die Relevanzrückkoppelung durchgeführt (vgl. Bildschirmausschnitt).

09	<input type="checkbox"/> ohne Titel Nationalrat: Sommersession 1999 Petition Forum der Tessiner Frauen für die Mutterschaftsversicherung Mutterschaftsversicherung Rechsteiner Paul (S, SG) 1159 Home URL: http://www.parlament.ch/Poly/Suchen_aml_Bulletin/cn99/ete/UD248.HTM?ervlet=get_content Subkollektionen: 1999, Sommersession, 1999, Summer Session, 1999, Session d'été, 1999, Sessione estiva
10	<input type="checkbox"/> Mutterschaftsversicherung Bundesgesetz / Assurance-maternité Loi fédérale Redner Geschäfte Orateurs Objets Nationalrat: Herbstsession 1998 Conseil national: Session d'automne 1998 97.055 Mutterschaftsversicherung. Bundesgesetz Assurance-maternité. Loi fédérale Vierzehnte Sitzung - Quatorzième séance Mittwoch, 7. Oktober 1998 - Me... URL: http://www.parlament.ch/Poly/Suchen_aml_Bulletin/cn98/automne/1057.HTM?ervlet=get_content Subkollektionen: 1998, Herbstsession, 1998, Autumn Session, 1998, Session d'automne, 1998, Sessione autunnale

[Relevanz-Rückkoppelung](#): Bitte kreuzen Sie relevante Dokumente an und klicken Sie auf den "Nochmals suchen"-Knopf

Manchmal kommt es vor, dass man während einer Recherche vom ursprünglich relevanten Themengebiet in einen Teilbereich oder in ein völlig anderes Gebiet gleitet. Dieses Problem nennen wir Abdriften.

Die Gefahr des ungewollten Abdriftens vom ursprünglichen Informationsbedürfnis während einer Relevanzrückkoppelung stellt sich vor allem bei Themen, die verschiedene Aspekte umfassen. Man sollte darauf achten, jeden Aspekt durch ein relevantes Dokument abzudecken.

Beispiel: Ein Ölscheich ist angsterfüllt auf der Suche nach Informationen über alternative Energiequellen. Er findet einiges über Solarenergie, Windenergie, Gezeitenkraftwerke und Ähnliches. Mit Hilfe einer Relevanzrückkoppelung möchte sich der Scheich weitere Dokumente beschaffen und sollte dazu möglichst zu jeder Art von Energiegewinnung ein Dokument einspeisen. Verwendet er lediglich einige Dokumente zum Thema Solarenergie, wird er in diesen Bereich abdriften und in erster Linie Resultate aus dem Gebiet der Solarenergie erhalten.



Der RotondoSpider wird als interne Suchmaschine des Schweizerischen Parlaments (www.parlament.ch 12.12.02) eingesetzt. Mache einige Tests zur Relevanzrückkoppelung. Als Suchanfrage kannst du Gebiete aus der Politik oder dem Studium wählen.

U1 Übungen

Frage 1

Wie heisst in der griechischen Mythologie der Hund, welcher den Eingang zu Unterwelt bewacht und jeden verschlingt, der daraus entfliehen will?

Welche körperliche Besonderheit besitzt dieser Hund?

Frage 2

Wie oft flog ein Schweizer Astronaut bereits ins All? Gesucht sind alle exakten Abflugsdaten und die Namen der Raumschiffe.

Frage 3

In der Verfassung der Eidgenossenschaft steht zu lesen: "Der Bund hat zum Zweck:, Handhabung von Ruhe und Ordnung im Innern, Schutz der Freiheit und der Rechte der Eidgenossen und Beförderung"

Aus welchem Artikel stammt der Text, und wie lauten die fehlenden Textstellen?

Frage 4

Erstbesteigung des Mount Everest:

- Beschaffe die offiziellen Daten: Jahr der Erstbesteigung, Namen und Nationalität der beteiligten Bergsteiger, Höhe des Bergs.
- Verschiedene Publikationen in den letzten 15 Jahren diese offizielle Version in Frage. Wann und von wem wäre der Everest gemäss dieser neuen These erstmals bestiegen worden, und durch welches Ereignis gewannen diesbezügliche Spekulationen im Frühling 1999 noch weiter an Aktualität?

Frage 5

Du solltest für deinen nächsten Deutsch-Vortrag unbedingt eine Biographie von Hermann Hesse beschaffen.

- Findest du im Internet etwas Passendes?
- Kannst du über Internet abklären, ob sich in der Zentralbibliothek Luzern ein geeignetes Buch über das Leben von Hesse finden liesse (um nicht eine vergebliche Fahrt nach Luzern zu machen)?

Frage 6

1995 wurde ein bekannter amerikanischer Footballstar und Schauspieler nach einer halsbrecherischen, von den TV-Kanälen live übertragenen Verfolgungsjagd verhaftet. Er wurde wegen Mordes an seiner Ex-Frau und deren Bekannten angeklagt.

- Wie hiess dieser Mann?
- Welches Auto (Marke und Typ) fuhr er bei der Verfolgung?
- Wie lautete schlussendlich der Urteilsspruch, und wann wurde er gefällt?

Frage 7

Im Jahre 1999 hat eine Gruppe von Computer- und Telekommunikationskonzernen eine neue, revolutionäre Technologie mit dem Namen „Blue Tooth“ der Öffentlichkeit vorgestellt.

- Wozu dient diese Technologie?
- Welche Firmen sind massgeblich an der Entwicklung beteiligt? (Nenne die 5 Gründungsmitglieder)
- Nach welcher historischen Figur ist die Technologie benannt?

U2 Übungen

Frage 1

Grosse Freude in der Schweiz und insbesondere an der ETH Zürich- Kurt Wüthrich erhält 2002 den Chemie-Nobelpreis. In der Geschichte des Nobelpreises hat es vor Kurt Wüthrich bereits diverse Schweizer Nobelpreisträger in den verschiedensten Sparten gegeben. Wer war der erste schweizer Nobelpreisträger? In welcher Sparte erhielt er den Nobelpreis?

Welche Schweizer Persönlichkeiten erhielten später ebenfalls diesen Nobelpreis? Wann?

Gab es Schweizer, die den Nobelpreis für Literatur gewonnen haben? Wenn ja wer und wann?

Frage 2

In welchem Lied kommt die Zeile „*there's a shadow hanging over me*“ vor? Wer war der Komponist und an welchem Tag wurde es erstmals aufgenommen?

Frage 3

Welches kriegerische Ereignis hielt vom 2.8.1990 bis zum 11.4.1991 die Welt in Atem? Was war der Anlass für dieses Ereignis? Welchen Codenamen hatte die Operation der Sieger? Wie hiess deren Kommandant?

Frage 4

Welcher bekannte Schweizer Maler malte 1922 das Bild „Herzdame“? Wo steht das Bild heute? Beschreibe kurz den Inhalt des Bildes.

Frage 5

Der Potsdamer Platz ist ein Wahrzeichen der deutschen Hauptstadt. Nenne mindestens 6 Strassen, die auf den Potsdamer Platz führen.

Frage 6

Es gibt in der Geschichte der römisch-katholischen Kirche seit dem Jahr 1500 gab es nur wenige Personen, die als Papst im gleichen Kalenderjahr gestorben sind, wie sie ihr Amt angetreten haben. Wie hiessen sie und in welchem Jahr amtierten sie als Pontifex Maximus?

Frage 7

In welcher Schweizer Stadt gibt es ein Schlossergässchen? Wo beginnt es und wo endet es (Strassennamen)?

Frage 8

Wer war 1972 Weltmeister in der Formel 1? Wie lautete sein Spitzname? Welches Auto fuhr er? Wieviele Siege erreichte er in der Saison 1972?

Frage 9

Welches radioaktive, chemische Element hat die Ordnungszahl 86? Welches Nuklid dieses Elements halbiert seine Radioaktivität in 55.61s?

Frage 10

Ende des 19. Jahrhunderts war der Beginn des eigentlichen Tourismus. In dieser Zeit wurden Hotelkästen aufgestellt und Naturgebiete erschlossen. Ebenfalls in dieser Zeit wurde die erste Zahnradbahn in Europa gebaut. Auf welchen Berg führte sie?

Wann genau wurde sie in Betrieb genommen?

Wer war der Erbauer und wann lebte er?