



Das Schweizer Portal für Recherche im Internet

sprint.informationswissenschaft.ch

A B C D E F **GLOSSAR** H I J K L M O P Q R S T U V W X Y Z

Alphabetisch

Entstanden im Rahmen des Projektkurses SPRINT,
überarbeitet im Herbstsemester 2008
an der Hochschule für Technik und Wirtschaft, HTW Chur

Stand: Januar 2009



Dieses Dokument untersteht der «Namensnennung-Keine kommerzielle Nutzung-Weitergabe unter gleichen Bedingungen 2.5 Schweiz»-Lizenz.

Was das bedeutet können Sie hier nachlesen: <http://creativecommons.org/licenses/by-nc-sa/2.5/ch/>

Zitiervorschlag:

Schweizer Portal für Recherche im Internet (SPRINT): Glossar alphabetisch. Chur: Hochschule für Technik und Wirtschaft, 2009. Im Netz: <http://sprint.informationswissenschaft.ch>

Glossar A

	Definition	Quelle
Abstandsoperatoren	Siehe Proximity Operators	
ACL (Agent-Communication-Languages)	Die Agent-Communication-Language dient zur Kommunikation zwischen den Agenten. Die ACL stellt ein standardisiertes Begriffssystem und eine formale Semantik für den Austausch von Nachrichten bereit. Damit sind Agenten nicht nur fähig Wissen und Informationen von anderen Agenten zu bekommen, sondern können auch gemeinsam an der Lösung einer Aufgabe arbeiten. Zur Kommunikation mit dem Benutzer dienen heute meist noch Dialogboxen. Der Einsatz natürlichsprachlicher Zugänge wird erforscht. Die bedeutendsten Kommunikationssprachen sind KQML (Knowledge Query Manipulation Language) und die ACL der Foundation for Intelligent Physical Agents (FIPA).	Tutorial Konstanz http://www.fh-wedel.de/~si/seminare/ws99/Ausarbeitung/agent/agent4.htm
Ajax	Ajax steht für „Asynchronous JavaScript and XML“. Es bezeichnet ein Konzept der asynchronen Datenübertragung zwischen einem Server und dem Browser, das es ermöglicht, innerhalb einer HTML-Seite eine HTTP-Anfrage durchzuführen, ohne die Seite komplett neu laden zu müssen. Das eigentliche Novum besteht in der Tatsache, dass nur gewisse Teile einer HTML-Seite oder auch reine Nutzdaten sukzessiv bei Bedarf nachgeladen werden, womit Ajax eine Schlüsseltechnik zur Realisierung des Web 2.0 darstellt.	http://de.wikipedia.org/wiki/Ajax_(Programmierung)
Archie-Server	Mit diesem Werkzeug lassen sich die FTP- Verzeichnisse mehrerer Server weltweit durchsuchen. Es können Datei- und Verzeichnisnamen gefunden werden. Es wird angegeben, wo welche Programme gefunden werden können.	http://www.suchfibel.de/4spez/spezialftp.htm
ASCII	ASCII ist ein Akronym für „American Standard Code for Information Interchange“ (dt.: Amerikanischer Standard-Code für den Informationsaustausch), der als ANSI-Standard X3.4 im Jahr 1968 eingeführt wurde. Als Vater des ASCII-Standards gilt Bob Bemer. ASCII beschreibt einen Zeichensatz, der auf dem lateinischen Alphabet basiert. Der Zeichensatz wird oft im modernen Englisch und von Computern und anderen Kommunikationseinrichtungen zur Textdarstellung verwendet. Er beschreibt als Code die Zuordnung von digital dargestellten Ganzzahlen zu den in der normalen Schriftsprache geschriebenen Zeichen. Mit Hilfe des Codes können digitale Geräte codierte Informationsinhalte senden, empfangen und verarbeiten. ASCII beschreibt einen Sieben-Bit-Code. Dieser Code verwendet binäre Ganzzahlen, die mit sieben binären Ziffern dargestellt werden, um Informationen darzustellen. Schon früh haben Computer mehr als 7 Bits, oft mindestens Acht-Bit-Zahlenworte, verwendet. Das achte Bit kann für Fehlerkorrekturzwecke (Paritätsbit) auf den Kommunikationsleitungen oder für andere Steuerungsaufgaben verwendet werden.	http://de.wikipedia.org/wiki/ASCII
Autoresponse-Mechanismus	Autoresponse-Mechanismus wird die automatische Weiterleitung von E-Mails genannt, die z.B. bei Abwesenheit oder für die Definition von Mail-	Tutorial Konstanz

	<p>Aliasnamen genutzt werden kann. Eine automatische Weiterleitung aller E-Mails an eine oder mehrere weitere Adressen lässt sich meist in Zusammenhang mit dem Mailserver (POP-Server) des für die betreffende Adresse zuständigen Mailsystems erreichen. Nützlich ist eine solche Einrichtung z.B. wenn man über mehrere Email-Adressen verfügt, die Mailbox jedoch nur von einer Stelle aus verwalten möchte.</p> <p>Eine automatisierte Beantwortung von E-Mails, die vor allem für längere Abwesenheit und ähnliche Situationen gedacht ist, kann auf ähnliche Weise eingerichtet werden.</p>	
--	--	--

Glossar B

	Definition	Quelle
Base64	<p>Base64 beschreibt ein Verfahren zur Kodierung von 8-Bit-Binärdateien. Es verwendet 65 druckbare ASCII-Zeichen des ISO 7-Bit-Codes, die auch über 7-bit-Verbindungen sicher übermittelt werden können. Dazu werden drei mal acht Bit zur Verschlüsselung von vier Zeichen mit je 6 Bit Datenbreite (denn $26 = 64$) „umfunktioniert“. Dem 65. Zeichen kommt eine Sonderfunktion zu. Das Verfahren findet im Internet-Standard MIME (Multipurpose Internet Mail Extensions) Anwendung und wird damit hauptsächlich zum Versenden von EMail-Anhängen verwendet. Nötig ist dies, um den problemlosen Transport von beliebigen Binärdaten zu gewährleisten, da SMTP in seiner ursprünglichen Fassung nur für den Versand von 7-Bit ASCII-Zeichen ausgelegt war. Bei der Kodierung von Binärdaten sollte auf Base64 als Kodierungsverfahren Wert gelegt werden, da die übertragenen Daten ansonsten für den Empfänger nicht nutzbar sein können. Nicht nutzbar heisst, dass die Daten unter Umständen vom lokalen Mail User Agent nicht mit ihrem korrekten Datentyp erkannt und entsprechend separiert werden können.</p>	<p>Tutorial Konstanz</p> <p>http://de.wikipedia.org/wiki/Base64</p>
Benutzerdefinierte Suchmaschine	<p>Mithilfe einer benutzerdefinierten Suchmaschine, welche von Google angeboten wird, kann die Suche auf lokale Server eingeschränkt werden. Der Nutzer kann ein Suchfeld und die Suchergebnisse auf seiner Webseite hosten. Dabei kann er selbst bestimmen, welche Webseiten bei einer Suchanfrage durchsucht werden sollen. Diese Dienstleistung kann von Unternehmen oder auch von privaten Personen, z.B. auf der eigenen Blogseite, genutzt werden.</p>	<p>http://www.google.com/coop/cse/</p>
BINARY	<p>Binaries sind mit UUEncode oder MIME in Binärdaten umgewandelte Dateien (Bilder, Daten, Software, ...), die somit per E-Mail oder in Newsgroups (Usenet) übertragen werden können.</p>	<p>http://de.wikipedia.org/wiki/Bin%C3%A4rdatei</p> <p>http://de.sybinf.org/b0000030.php</p>
Bitnet	<p>Das Bitnet (Akronym von Because It`s Time NETWORK) war ein akademisches Netzwerk, das über 3000 Host-Rechner in mehr als 32 Ländern miteinander verband. Das Bitnet entstand an der City University of New York und war mit identischen Netzen in Kanada und Europa mit den Namen NetNorth und EARN verbunden. Neben der akademischen Zusammenarbeit bot das Bitnet auch</p>	<p>http://de.wikipedia.org/wiki/BITNET</p>

	Diskussionsforen, E-Mail-Dienste, FTP (BitFTP) und vieles anderes mehr. Auf seinem Höhepunkt zu Beginn der 1990er Jahre kommunizierten an die 3500 Rechner in über 1400 Organisationen miteinander. Danach wurden die Rechner immer kleiner und zahlreicher. Das Internet übernahm die Kommunikation und das Bitnet wurde aufgelöst.	
Blog	Siehe Weblog	
Blogosphäre	Der Begriff Blogosphäre (engl. „blogosphere“) beschreibt die Gesamtheit der Weblogs und ihrer Verbindungen. Er entspringt der Wahrnehmung, dass Blogs durch ihre Vernetzungen gemeinsam eine oder eine Vielzahl von Communities bilden, beziehungsweise ein soziales Netzwerk darstellen.	http://de.wikipedia.org/wiki/Blogosph%C3%A4re
Bookmark	Im Internet versteht man unter Bookmark eine Linksammlung in Anwenderprogrammen, meist Webbrowser. Je nach Anwenderprogramm werden unterschiedliche Bezeichnungen verwendet, z.B. „Lesezeichen“ im Firefox oder „Favoriten“ im Internet Explorer. Bei englischsprachigen Programmen findet sich häufig die Bezeichnung Bookmark. Lesezeichen helfen, eine einmal gefundene Seite auch ohne Suchmaschine wieder zu finden. Erfahrene Internetanwender sammeln Lesezeichen und kommen so häufig auf tausende von Einträgen. Viele Webbrowser, darunter Mozilla und Linux, verwalten Lesezeichen in einer HTML-Datei (meistens kein reines HTML, sondern HTML mit speziellen Erweiterungen). Opera verwendet eine speziell formatierte Textdatei. Bei den Favoriten des Internet Explorers wird pro Eintrag eine kleine Datei real im Dateisystem abgelegt. Dies geschieht strukturiert, in Ordnern.	http://de.wikipedia.org/wiki/Lesezeichen_(Internet)
Boolesche Operatoren	Boolesche Operatoren sind logische Operatoren, die auf der booleschen Algebra (benannt nach Georg Boole) beruhen. Boolesche Operatoren beinhalten Ausdrücke wie UND, ODER, NICHT und XOR (ausschliessendes ODER).	http://de.wikipedia.org/wiki/Boolescher_Operator
Breadth-first	Breadth-first search (deutsch: Breitensuche) ist ein Fachbegriff der Informatik, welcher ein Verfahren zum Durchsuchen beziehungsweise Durchlaufen der Knoten eines Graphen bezeichnet. Die Suchroboter gehen dabei alle abgehenden Verweise (Links) einer Seite nach und laden die so erreichten Zieldokumente herunter. Es wird erst dann ein weiterer Tiefenschritt in der Hyperlink-Verfolgung vorgenommen, wenn bereits alle anderen Links erschlossen sind. Breitensuche steht im Gegensatz zur depth-first-search/Tiefensuche.	http://de.wikipedia.org/wiki/Breitensuche
Browsing	Zielgerichtete oder auch weniger zielgerichtete Navigation von einem Hyperlink zum anderen.	Tutorial Konstanz
B-to-B	B-to-B (auch B2B) ist die Abkürzung für Business-To-Business und steht für (elektronische) Kommunikationsbeziehungen zwischen Unternehmen. Im Gegensatz dazu steht die Kommunikation von Unternehmen mit Privatpersonen, mit Mitarbeitern oder mit der öffentlichen Verwaltung.	http://de.wikipedia.org/wiki/Business-To-Business
Business Intelligence	Der Begriff Business Intelligence bezeichnet Verfahren und Prozesse zur systematischen Analyse (Sammlung, Auswertung und Darstellung) von Daten in elektronischer Form. Ziel ist die Gewinnung von	http://de.wikipedia.org/wiki/Business_Intelligence

	Erkenntnissen, die in Hinsicht auf die Unternehmensziele bessere operative oder strategische Entscheidungen ermöglichen. Dies geschieht mit Hilfe analytischer Konzepte und IT-Systeme, die Daten über das eigene Unternehmen, die Mitbewerber oder Marktentwicklung im Hinblick auf den gewünschten Erkenntnisgewinn auswerten. Mit den gewonnenen Erkenntnissen können Unternehmen ihre Geschäftsabläufe, Kunden- und Lieferantenbeziehungen profitabler machen, Kosten senken, Risiken minimieren und die Wertschöpfung vergrößern.	
--	--	--

Glossar C

	Definition	Quelle
Caching	Beim Caching werden bereits abgerufene Seiten so zwischengespeichert, dass sie im Fall eines erneuten Abrufs nicht noch einmal von der ursprünglichen URL geladen werden müssen. Bei diesem schnelleren Zugriff werden die Daten vom Zwischenspeicherort herunter geladen. Caching kann auch durch die Speicherung von abgerufenen Dokumenten im Hauptspeicher oder auf der Festplatte des lokalen Rechners erfolgen. Es kann auch in Form von so genannten Proxy-Servern vom Informations-Provider bereitgestellt werden.	Tutorial Konstanz
Case-Sensitivity	Case-Sensitivity heisst, dass zwischen Klein- und Grossschreibung der Buchstaben unterschieden wird. Die meisten Suchdienste sind jedoch nicht case-sensitive.	http://de.wikipedia.org/wiki/Case_sensitivity
CGI (Common Gateway Interface)	Bei CGI, dem Common Gateway Interface, handelt es sich um eine serverseitige Technologie zur Verbindung der Web-Clients mit Anwendungen und Daten, die auf dem Server lokalisiert sind. Ermöglicht wird das Übergeben von Daten an Server bspw. zur Erfassung und Verarbeitung von Einträgen, die der Nutzer in HTML-Formularfelder vorgenommen hat. Mit CGI wird eine standardisierte Schnittstelle definiert, die über Umgebungsvariablen genutzt werden kann. Die Funktionalität wird dabei auf dem Server in Form von Programmskripten umgesetzt. Damit solche Skripte auf dem Server Verwendung finden können, muss dieser CGI unterstützen. CGI ist unabhängig von einer bestimmten Programmiersprache, häufig aber wird Perl eingesetzt, weil dieses bei Unix-basierten Servern in Form eines bereitgestellten Interpreter-Programms, als Voraussetzung für die Abarbeitung von CGI-Skripten, meistens installiert ist. Die Einbindung der CGI-Skripte erfolgt in HTML über die URL der Datei mit dem Skript-Programm.	Tutorial Konstanz
Cloaking	Der Begriff „Cloaking“ stammt aus dem englischen und bedeutet eigentlich verhüllen. Im Zusammenhang mit der Suche im Internet spricht man von Cloaking, wenn zwei oder mehr verschiedene Versionen einer Homepage existieren. Das bedeutet, dass eine oder mehrere Seiten für die Roboter der Suchmaschinen optimiert sind und dass gleichzeitig eine „normale Version“ für die Benutzer vorliegt. Die Version für die Suchmaschinen dient dazu, eine bessere Platzierung im	Tutorial Konstanz

	<p>Ranking zu erzielen und kann beispielsweise mit relevanten Suchbegriffen versehen sein. Auf gestalterische Elemente wie Frames usw. wird verzichtet. Wenn die Ranking-Algorithmen der Suchmaschinen bekannt sind, kann für jede Suchmaschine eine eigene, optimierte Version erstellt werden. Dieses Verfahren funktioniert, weil sich Suchmaschinen bei der Indexierung entgegenkommenderweise zu erkennen geben, um dem Homepage-Anbieter zu ermöglichen, die Anfragen von Suchdiensten zurückzustellen und zuerst die Anfragen der regulären Besucher zu bearbeiten.</p>	
Cluster	Siehe Clustering	
Clustering	<p>Ein viel verbreiteter Mechanismus bei der Visualisierung der Trefferanzeige ist das Clustering. Suchmaschinen mit Clustervisualisierung unterteilen die ermittelte Ergebnismenge in unterschiedliche grobe Kategorien und erleichtern dem Nutzer dadurch die Selektion der für ihn relevanten Trefferbereiche. Die Clusterbildung spielt im Information Retrieval allgemein eine sehr wichtige Rolle, da hierbei Dokumente, so ähnlich wie dies bei Web-Katalogen der Fall ist, nach gemeinsamen inhaltlichen Kriterien gruppiert werden und dem Benutzer auf diese Weise die sonst bei Suchmaschinen fehlenden Kontextinformationen zur Verfügung stellen.</p>	<p>Weinhold, T. / Bekavac, B. / Hierl, S. / Öttl, S. (2008): Visualisierung bei Internetsuchdiensten. In: Dirk Lewandowski (Hrsg.), Handbuch Internet-Suchmaschinen - Nutzerorientierung in Wissenschaft und Praxis. Heidelberg: Akademische Verlagsgesellschaft.</p>
CMS (Content Management System)	<p>Ein Content Management System (CMS) ist eine Software, die die Verwaltung der Inhalte von Webseiten ermöglicht, wobei eine Trennung zwischen Inhalt, Gestaltung und Funktion möglich ist. CMS ermöglichen Benutzern die Erstellung und die Verwaltung von Webinhalten auch ohne Programmierkenntnisse. Weitere Vorteile von CMS sind die Möglichkeit zur dezentralisierten Wartung (d.h. der Zugriff auf das System ist über jeden Webbrowser möglich), die Erhaltung von Design-Vorgaben (durch die Trennung von Inhalt und Layout), die Speicherung der Inhalte in einer zentralen Datenbank (erleichtert die Wiederverwendung von Inhalten), sowie die automatische Generierung der Navigations-Menüs (diese werden typischerweise aus den Datenbankinhalten erzeugt). CMS bieten zudem eigene, interne Suchfunktionen an. Die Suche kann dabei auch auf Teile einer Website beschränkt werden und sogar die Suche in Datenbank-Tabellen ist möglich. Bei CMS wird keine CGI-Schnittstelle mehr benötigt.</p>	<p>http://de.wikipedia.org/wiki/Content-Management-System</p>
Crawler	<p>Ein Crawler (auch Spider) ist ein Computerprogramm, welches automatisch das WWW durchsucht und Webseiten analysiert. Crawler werden vor allem von Suchmaschinen eingesetzt. Weitere Anwendungen sind beispielsweise das Sammeln von EMailadressen. Ein Crawler gelangt über Hyperlinks von einer Webseite zu weiteren URLs. Dabei werden alle bisher bekannten Adressen gespeichert und der Reihe nach besucht. Die neu gefundenen Hyperlinks werden zur Liste aller URLs hinzugefügt. So können theoretisch alle erreichbaren Seiten des WWW gefunden werden. Je nach Aufgabe des Crawlers wird der Inhalt der gefundenen Webseiten bspw. mittels Indexierung ausgewertet und gespeichert, um ein späteres Suchen</p>	<p>http://de.wikipedia.org/wiki/Webcrawler</p>

	in den gesammelten Daten zu ermöglichen.	
CSS (Cascading Style Sheets)	Cascading Style Sheets (CSS) ist eine deklarative Stylesheet-Sprache für strukturierte Dokumente (z.B. HTML und XML). Durch die Trennung von Stil und Inhalt wird das Veröffentlichen und Betreuen von Dokumenten wesentlich vereinfacht. CSS wurde vor allem im Hinblick auf HTML entwickelt, ist aber auch für XML-Dokumente anwendbar. CSS ermöglicht es auch, Inhalte nach dem Ausgabemedium (z.B. Druck, Handy) zu ändern. Das ist nützlich, um zum Beispiel Weblinks beim Drucken extra aufzuführen und nicht (wie oft bei HTML-Seiten) zu verbergen.	http://de.wikipedia.org/wiki/Cascading_Style_Sheets

Glossar D

	Definition	Quelle
Dead-Link	Ein Dead-Link ist ein toter, also nicht mehr aufrufbarer Hyperlink. Dead-Links entstehen beispielsweise, wenn das Zieldokument umbenannt, gelöscht oder die URL falsch eingegeben wurde. Der Webserver liefert dann normalerweise die Fehlermeldung „404 - File not found“.	http://de.wikipedia.org/wiki/Toter_Link
Deep Web	Siehe Invisible Web	
Depth-First	Depth-first search (deutsch: Tiefensuche) ist ein Fachbegriff der Informatik, welcher ein Verfahren zum Durchsuchen beziehungsweise Durchlaufen der Knoten eines Graphen bezeichnet. Bei der Suche werden dabei die Links über mehrere Seiten hinweg verfolgt, das heisst, die Knoten soweit verfolgt, bis der letzte Knoten erreicht ist. Danach geht die Suche zurück, um von einem höher gelegenen Knoten wieder abzusteigen. Tiefensuche steht im Gegensatz zur breadthfirst-search/Breitensuche.	http://de.wikipedia.org/wiki/Tiefensuche
Digitale Kluft / digital divide	Unter dem Schlagwort „digital divide“ oder „digitale Kluft“ fand in den zurückliegenden Jahren eine umfangreiche öffentliche Auseinandersetzung um den Einfluss der neuen Informations- und Kommunikationstechnologien auf die Verteilung gesellschaftlicher Ressourcen und somit auf die Gewährleistung von Chancengleichheit und sozialer Gerechtigkeit in der Informationsgesellschaft statt. Ausgangspunkt der Diskussion ist die empirisch beobachtbare Tatsache, dass diese Technologien, und dabei insbesondere das Internet, sich zwar mit grosser Geschwindigkeit ausbreiten, diese Ausbreitung allerdings sozial sehr heterogen erfolgt. Die sich daraus ergebenden Differenzen im Technologiezugang verschiedener sozialer Gruppen werden als Zugangsklüfte bezeichnet. Problematisch an diesen Zugangsklüften ist, dass vor allem solche sozialen Gruppen von den Anwendungsmöglichkeiten von Computer und Internet ausgeschlossen (unterrepräsentiert) bleiben, die ohnehin schon zu den unterprivilegierten Schichten der Gesellschaft gehören. Aus der Deutung dieser Zugangsungleichheiten als Vorboten einer neuen Zweiklassengesellschaft, mit einer technologie-versierten Informationselite auf der einen Seite und einer technologie-abstinenten Informationsparia auf der	http://beat.doebe.li/bibliothek/w00870.html

	<p>anderen Seite, ergibt sich ein Bedrohungsszenario, dem mit zahlreichen Fördermassnahmen begegnet wird. Generelles Ziel dieser Massnahmen ist die Gewährleistung eines Internetzugangs für alle Bürgerinnen und Bürger.</p>	
<p>DNS (Domain Name System)</p>	<p>Das Domain Name System (DNS) ist einer der wichtigsten Dienste im Internet. Das DNS ist eine verteilte Datenbank, die den Namensraum im Internet verwaltet.</p> <p>Das DNS wird hauptsächlich zur Umsetzung von Namen in Adressen benutzt. Dies ist vergleichbar mit einem Telefonbuch, welches die Namen der Teilnehmer in ihre Telefonnummer auflöst. Das DNS bietet somit eine Vereinfachung, da Domainnamen wie z.B. google.com einfacher merkbar sind als die dazugehörige IP-Adresse (z.B. 111.222.333.444). Mit dem DNS ist auch eine umgekehrte Auflösung von IP-Adressen in Namen möglich.</p> <p>Das DNS besteht aus den drei Komponenten Domänennamensraum, Nameserver und Resolver. Ein konkretes Beispiel für einen Domänennamen ist google.com.</p> <p>Nameserver sind Programme, die einen oder mehrere Teile des Namensraumes kennen und diese auf Anfrage weitergeben.</p> <p>Resolver sind Programme, die Informationen aus den Nameservern abrufen können. Sie bilden die Schnittstelle zum Nameserverdienst. Resolver sind entweder eigene Programme, oder sie sind in Applikationen (z.B. einen Browser) eingebunden.</p>	<p>http://de.wikipedia.org/wiki/Domain_Name_System</p>
<p>Domain</p>	<p>Eine Domäne (engl. Domain) ist ein zusammenhängender Teilbereich des hierarchischen DNS Namensraums. Eine Domäne umfasst ausgehend von ihrem Domainname immer die gesamte untergeordnete Baumstruktur.</p> <p>Beispiel: Gegeben sei die Domäne google.com. Es existieren weiterhin ein Host www.google.com sowie z.B. die Subdomain (Domänen unterhalb von Domänen) news.google.com. Alles diese Knoten gehören zur Domäne google.com.</p> <p>Domänen der höchsten Ebene (z.B. .com) werden auch Top-Level-Domains genannt. Dementsprechend heissen die Domänen zweiter und dritter Ebene second- bzw. third-leveldomains.</p>	<p>http://de.wikipedia.org/wiki/Domain</p>
<p>DTD (Document Type Definition)</p>	<p>Die Dokumenttypdefinition ist eine Deklaration in SGML- und XML-Dokumenten, die die Struktur eines solchen Dokuments festlegt. Eine DTD beschreibt beispielsweise HTML oder XHTML. DTDs sind auch in speziellen Einsatzgebieten verbreitet wie z.B. DocBook.</p> <p>In einer DTD werden Elemente, Attribute von Elementen und Besonderheiten der Syntax-Verwendung definiert. Letzteres ist jedoch nur bei bestimmten SGML-Deklarationen möglich. Dies heisst konkret, dass in einer DTD beispielsweise die Reihenfolge der Elemente oder die Art des Inhalts von Attributen festgelegt wird/werden kann.</p> <p>Die Syntax und Semantik einer DTD ist Bestandteil des XML-Standards. Die DTD wird zu Beginn eines XML-Dokumentes in der Dokumenttyp-Deklaration angegeben. Die Grammatikregeln der DTD können dabei sowohl innerhalb des XML-Dokumentes (interne</p>	<p>http://de.wikipedia.org/wiki/Dokumenttypdefinition</p>

	DTD) als auch in einer externen Datei angegeben werden.	
Dublin Core Metadata Element Set	<p>Der Dublin Core wurde zwischen 1995 und 1996 entwickelt um ein besseres Retrieval von Informationen, insbesondere im World Wide Web, zu erhalten. Benannt wurde der Dublin Core nach dem Ort der ersten Tagung, Dublin, Ohio. Dublin Core wurde als genereller Standard für Online-Informationen entwickelt, der das Beschreiben von HTML-, PDF- und Grafikobjekten erlaubt. Dublin Core ist deskriptiv und nicht evaluierend. Deswegen eignet sich Dublin Core nicht besonders gut für die Bewertung von Dokumenten. Bei der Standardisierung von Dublin Core wurde besonderer Wert auf Einfachheit gelegt, so dass die Elementmenge auf einen Kern von 15 Elementen zur inhaltlichen und formalen Erschliessung, begrenzt wurde.</p> <p>Beispiele von Elementen:</p> <ul style="list-style-type: none"> - TITLE (Name der Quelle) - CREATOR (Name der für den intellektuellen Inhalt der Quelle verantwortlichen Person/en) - DESCRIPTION (Beschreibung des Inhalts der Quelle) - DATE (Zeitpunkt, zu dem die Quelle in ihrer jetzigen Form zugänglich gemacht wurde) 	http://de.wikipedia.org/wiki/Dublin_Core

Glossar E

	Definition	Quelle
Envelope	<p>Envelope ist neben Header und Body einer der drei wichtigen Bausteine, die für eine Übertragung einer Mail im Internet wichtig sind. Im Envelope stehen Steuerungsdaten für die Mailserver, um die Nachricht korrekt zu übermitteln. Der Anwender sieht diese Daten in der Regel nie, da der Mailserver beim Ablegen der Mail im Postfach diese Daten löscht.</p> <p>Die Informationen über Absender (FROM) und Empfänger (TO) im Header dienen der Anzeige beim Anwender. Die Informationen über Absender (ENVELOPE MAIL FROM) und Empfänger (ENVELOPE MAIL TO) im Envelope dienen den Mailservern (SMTP-Servern) dazu, die Nachrichten zu routen, d.h. korrekt weiterzuleiten. Im Body befinden sich der Mailinhalt und eventuelle Mailanlagen.</p>	http://www.msxch.angefaq.de/internet/envelope.htm

Glossar F

	Definition	Quelle
FAQ (Frequently Asked Questions)	<p>Bei den FAQ (Frequently Asked Questions) handelt sich um eine Zusammenstellung von oft gestellten Fragen und den dazugehörigen Antworten zu einem Thema. Bekannt geworden sind FAQs im IT-Bereich, insbesondere im Internet.</p> <p>In vielen NewsGroups gibt es eine so genannte FAQ-Datei. Diese Dateien beinhalten eine Sammlung der in der jeweiligen Gruppe am Häufigsten gestellten Fragen und den dazugehörigen Antworten. Bevor ein Benutzer eine Frage an eine NewsGroup stellt, sollte er diese</p>	<p>Tutorial Konstanz</p> <p>http://de.wikipedia.org/wiki/FAQ</p>

	<p>FAQ-Datei lesen. So werden die regelmässigen Leser einer Gruppe nicht immer wieder mit den gleichen Fragen konfrontiert.</p> <p>Der Standort der FAQ-Dateien wird in regelmässigen Abständen in der Gruppe veröffentlicht. FAQ-Dokumente sind im Internet aber auch zu vielen anderen Themen verfügbar, unabhängig von entsprechenden NewsGroups.</p>	
Firewall	<p>Als Firewall oder Zugangsschutzsystem bezeichnet man bei Rechnernetzwerken ein organisatorisches und technisches Konzept zur Trennung von Netzbereichen, dessen korrekt Umsetzung und dauerhafte Pflege. Durch den immer grösser werdenden Ausbau von Netzen wird der Schutz einzelner Netze immer wichtiger. Firewalls greifen hier ein. Sie sitzen an den Schnittstellen zwischen einzelnen Netzen und kontrollieren den Netzwerkverkehr zwischen den Netzen, um unerwünschten Verkehr zu verhindern und nur den gewünschten Verkehr weiterzuleiten. Der häufigste Einsatz einer Firewall besteht darin, den Verkehr zwischen einem lokalen Netzwerk und dem Internet zu kontrollieren und zu steuern. Umgangssprachlich ist mit einer Firewall sehr oft die Software gemeint, welche den Datenverkehr zwischen den getrennten Netzbereichen kontrolliert und regelt. Es muss also zwischen dem (Sicherheits-)Konzept Firewall und den zwei Hauptbestandteilen der Firewall, nämlich Hardware und Software, unterschieden werden. Die Hardware ist für das Empfangen und Senden der einzelnen Netzwerkpakete zuständig und die Software regelt den Verkehr.</p>	http://de.wikipedia.org/wiki/Firewall
Flash	<p>Adobe Flash (kurz Flash) ist eine proprietäre integrierte Entwicklungsumgebung zur Erstellung multimedialer Inhalte, der Flash-Filme. Die resultierenden Dateien liegen im SWF-Format vor, einem auf Vektorgrafiken basierenden Grafik- und Animationsformat. Das Kürzel SWF steht dabei für Shockwave Flash.</p>	http://de.wikipedia.org/wiki/Adobe_Flash
Flex	<p>Adobe Flex (kurz Flex) ist ein Entwicklungsframework zum Erstellen von Rich Internet Applications (RIAs). RIA steht für eine Anwendung, die Internet-Techniken benutzt und eine intuitive Benutzeroberfläche bietet. Schon vor der Veröffentlichung von Flex war es möglich mit Adobe Flash RIAs zu erstellen, viele Entwickler schauten jedoch darüber hinweg, wurde Flash doch immer als Werkzeug für Webdesigner und Grafiker gesehen. Adobe versucht Flex als eine Plattform zu etablieren, die auch Softwareentwickler anspricht.</p>	http://de.wikipedia.org/wiki/Adobe_Flex
Folksonomy	<p>Der Begriff Folksonomy setzt sich aus den englischen Wörtern „Folks“ und „Taxonomy“ zusammen. Unter Folksonomy wird eine Taxonomie, also Systematik, verstanden, die von den Computer-Nutzern festgelegt und aufrechterhalten wird. Die Systematik wird dabei nicht a priori festgelegt, sondern ergibt sich organisch. Folksonomien entstehen im Web mit Hilfe von Social Software Anwendungen wie z.B. die Bilderdatenbank Flickr.</p>	http://www.e-teaching.org/glossar/folksonomy
Frame	<p>Mit Hilfe von Frames kann der Anzeigebereich des Browsers in verschiedene, frei definierbare Segmente aufgeteilt werden, das heisst, es können verschiedene Informationen in einem Browserfenster dargestellt werden. Dazu wird eine Technik benötigt, um das</p>	http://de.wikipedia.org/wiki/Frame_(HTML)

	<p>Fenster selbst unterteilen zu können. Dies geschieht mit einem sogenannten frameset. Mit dieser HTML Anweisung legt man fest, wie das Browserfenster unterteilt werden soll: waagrecht oder senkrecht sowie die Anzahl der Segmente. Die hierbei entstandenen Bereiche nennt man Rahmen (Frames). Für jeden der entstandenen Frames kann beziehungsweise muss festgelegt werden, was in diesem (anfänglich) dargestellt werden soll. Jedes Segment kann eigene Inhalte enthalten. Die einzelnen Anzeigesegmente (Frames) können einen statischen Inhalt oder einen wechselnden Inhalt haben. Verweise in einem Frame können Dateien aufrufen, welche in einem zusätzlichen Frame angezeigt werden.</p>	
<p>FTP (File Transfer Protocol)</p>	<p>Das File Transfer Protocol (engl. für Dateiübertragungsverfahren) ist ein Netzwerkprotokoll zur Dateiübertragung über TCP/IP-Netzwerke. FTP ist in der Anwendungsschicht des TCP/IP-Protokollstapels angesiedelt. Es wird benutzt, um Dateien vom Server zum Client (Download), vom Client zum Server (Upload) oder clientgesteuert zwischen zwei Servern zu übertragen.</p> <p>Das signifikanteste Merkmal von FTP ist die Trennung von Daten- und Kontrollfluss. Über die Kontrollverbindung werden zunächst die Parameter der Datenübertragung wie Datentyp, Übertragungsmodus, durchzuführende Operationen usw. festgelegt. Erst wenn dies erfolgt ist, beginnt der eigentliche Datentransfer über die Datenverbindung. Sowohl der Verbindungsaufbau als auch die Steuerkommandos werden über die Kontrollverbindung realisiert.</p> <p>Der FTP-Dienst wurde einst zur Übertragung grosser Datenmengen entwickelt. Dieser Dienst existierte schon lange vor dem WWW. Zum Zugriff auf FTP-Server sind eine Benutzerkennung und ein zugehöriges Passwort nötig. Das FTP-Verfahren kommt im Web meist beim Download von Daten zum tragen. Häufig bemerken Benutzer gar nicht, dass zum Zeitpunkt der Übertragung von Dateien von HTTP auf FTP gewechselt wird. Beim Zugang zu FTP-Verzeichnissen über einen Web-Browser wird die Benutzeranmeldung vom Browser automatisch vorgenommen und das FTP-Dateiverzeichnis wird im „Look and Feel“ des WWW angezeigt. Das Spektrum der verfügbaren Daten auf FTP-Servern ist sehr breit gefächert. Beispiele für diese Daten sind z.B. Public Domain, Freeware und Shareware für alle Rechnertypen, Treiber, Updates, Forschungsberichte, FAQ-Dateien, Videodateien etc.</p>	<p>Tutorial Konstanz</p> <p>http://de.wikipedia.org/wiki/File_Transfer_Protocol</p>
<p>FTP-Server, öffentliche</p>	<p>Auf öffentlichen FTP-Servern werden diverse Daten und Softwares zum Download angeboten. Bei öffentlichen FTP-Servern gibt es allgemein bekannte und gültige Benutzerkennungen wie z. B. „guest“, „ftp“ oder „anonymous“ sowie als Passwort eine oder die eigene E-Mail-Adresse. Im Gegensatz zu öffentlichen FTP-Servern sind bei nicht-öffentlichen FTP-Servern Benutzerberechtigungen definiert, der Login erfolgt über einen Benutzernamen sowie ein persönliches Passwort.</p>	<p>Tutorial Konstanz</p> <p>http://de.wikipedia.org/wiki/File_Transfer_Protocol</p>

Glossar G

	Definition	Quelle
Gateway	Ein Gateway erlaubt Netzwerken, die auf völlig unterschiedlichen Protokollen basieren, miteinander zu kommunizieren. Zu diesem Zweck nimmt ein Gateway eine Protokollumsetzung vor. Dem Gateway ist dabei alles erlaubt, was zur Konvertierung der Daten notwendig ist, auch das Weglassen von Informationen, wenn diese im Zielnetz nicht transportiert werden können. Im Detail werden sämtliche Protokollinformationen, die an ein Datenpaket angehängt werden, entfernt und durch andere ersetzt. Daneben gibt es auch Gateways für zahlreiche andere Verwendungszwecke wie etwa E-Mail zu SMS, Fax zu E-Mail etc.	http://de.wikipedia.org/wiki/Gateway_(Computer)
Gewichtung	Durch die Gewichtung der Suchbegriffe bei einer Suchmaschinen-Anfrage kann die Treffermenge eingeschränkt bzw. die Menge an relevanten Treffern erhöht werden. Stärker gewichtete Suchbegriffe werden dann bei der Suche bevorzugt. Dabei variiert die Art der Angabe nach Häufigkeit des Auftretens eines Begriffs in der Suchanfrage bis zur Angabe eines Gewichtungsgrades. Auch unterschiedliche Stellen des Dokumentes an denen die Suchbegriffe vorkommen, können flexibel gewichtet werden. Google beispielsweise beachtet die Reihenfolge der Begriffe. Wichtige Begriffe sollten deshalb zuerst eingegeben werden.	http://www.google.de/support/bin/static.py?page=searchguides.html&ctx=basics
GIF (Graphics Interchange Format)	GIF ist die Abkürzung für „Graphics Interchange Format“. GIF ist ein digitales Bildformat mit guter, verlustfreier Komprimierung für Bilder mit geringer Farbtiefe sowie mit der Möglichkeit, einen transparenten Hintergrund zu nutzen und animierte Bilder zu erstellen. GIF wurde 1987 vom US-Onlinedienst CompuServe eingeführt und wurde vor allem wegen seiner effizienten Kompression populär.	http://de.wikipedia.org/wiki/GIF
Gopher	Gopher (engl. Erdhörnchen) ist ein Informationsdienst, der über das Internet mit Hilfe eines Gopher-Clients oder einem Webbrowser abgerufen werden kann. Gopher hat Ähnlichkeit mit dem frühen WWW und wurde 1992 an der Universität von Minnesota entwickelt. Gopher entstand also früher als das WWW und hat deswegen auch eine geringfügig niedrigere Portnummer (70 - im Vergleich zu 80 für http). Die Überlegung, die zu Gopher führte, war die umständliche Handhabung des FTP, bei dem man sich einloggen und über Konsolenbefehle in Verzeichnisse wechseln musste, um die gewünschte Datei finden und herunterladen zu können. Zudem wollte man ein einfach zu administrierendes Informationssystem schaffen, welches wenige Ressourcen benötigt. Mitte der 90er Jahre hatte nahezu jede Organisation, die über einen Internetzugang verfügte (Universitäten, Regierungen etc.), einen Gopherserver, und stellte der Allgemeinheit Informationen aus allen Bereichen zur Verfügung. Mit dem Aufschwung des WWW ging jedoch die Zeit des sogenannten Gopherspace zu Ende. Heute gibt es nur noch sehr wenige Gopherserver, Gopher hat weitgehend an Bekanntheit verloren. Gopher baut auf das Gopher-Protokoll auf und bietet im	http://de.wikipedia.org/wiki/Gopher

	Gegensatz zu HTML-Seiten ein automatisch generiertes Menü an, das aus den im aktuellen Verzeichnis befindlichen Dateien generiert wird. Der Gopher-Server erkennt dabei, ob es sich um Verzeichnisse oder Dateien handelt.	
--	--	--

Glossar H

	Definition	Quelle
Header (E-Mail)	Im Header befinden sich die Adress- und Metainformationen einer E-Mail. Die Header-Zeilen werden durch Felder repräsentiert: Feldname: Inhalt <cr>. Kernbestandteil vieler Felder sind E-Mail-Adressen. Dazu gehören Absenderinformationen, Empfängerinformationen, inhaltsbezogene Angaben (z.B. Datum, Subject), Pfadangaben und E-Mail-Identifikation (Antwortpfad, Kennung der MTAs, eindeutige ID der E-Mail etc.) sowie (je nach Anwendung) auch frei wählbare Felder.	http://de.wikipedia.org/wiki/Header_(E-Mail)
HTML (Hypertext Markup Language)	HTML (Hypertext Markup Language) ist ein Layout-unabhängiges Dokumentbeschreibungsformat für Webseiten, das Hypertext-Verweise auf der Basis von URL enthält. HTML wurde 1989 von Tim Berners-Lee am CERN in Genf festgelegt. HTML basiert auf dem Dokumentenbeschreibungsstandard SGML, der zur Definition von Auszeichnungssprachen verwendet wird. Auszeichnungssprachen (Markup Languages) dienen zur Beschreibung von Daten oder des Verfahrens/der Schritte, die zur Darstellung nötig sind. HTML ist also eine Auszeichnungssprache zur Beschreibung von Informationen in Hypertexten. Die Hypertext-Elemente dienen zum Verweis auf andere Textstellen oder auf ein anderes Dokument. In HTML geht es nicht um die optische Gestaltung eines Textes, sondern viel mehr um die Zuordnung von Bedeutungen zu einzelnen Textbereichen (z.B. Überschriften). Ob ein bestimmter Textbereich durch Kursiv-, Fettschrift oder eine auffällige Farbe realisiert wird, ist zweitrangig und dem Webbrowser überlassen, kann aber mit Hilfe von CSS beeinflusst werden. HTML-Dokumente sind in 3 Bereiche aufgeteilt: <ul style="list-style-type: none"> - Information über die verwendete HTML-Version - Ein Kopfteil mit Metainformationen zum Dokument (<HEAD>) - Ein Rumpfteil mit dem eigentlichen Inhalt (<BODY>) Die restlichen Elemente dienen hauptsächlich zur Gestaltung (Hyperlinks, Überschriften, Bilder, Frames etc.).	http://de.wikipedia.org/wiki/HTML
http (Hypertext Transfer Protocol)	Das Hypertext Transfer Protocol (http) ist ein zustandsloses, ASCII-basiertes Datenaustausch-Protokoll zur Datenübertragung. Primär wird es zur Übertragung von Webseiten verwendet (Web-Browser greifen fast ausschliesslich mit diesem Protokoll auf Web-Server zu). Das Protokoll wurde 1989 von Tim Berners-Lee am CERN in Genf zusammen mit URL und HTML erfunden. Dies war gleichzeitig die Geburtsstunde des WWW. HTTP liegt das Prinzip der Datenübermittlung nach dem	http://de.wikipedia.org/wiki/Hypertext_Transfer_Protocol http://www.tecchannel.de/netzwerk/management/401210/hypertext_transfer_protocol/index.html

	Request-Response-Schema zugrunde. Der HTTP-Client sendet seine Anfrage an den HTTP-Server, der diese bearbeitet und eine Antwort zurücksendet. Ein Vorteil von HTTP ist, dass es grundsätzlich abwärtskompatibel ist. So können zum Beispiel Browser, die HTTP 1.1 beherrschen, auch mit HTTP-1.0-kompatiblen Webservern kommunizieren. Die jeweils höhere Versionsnummer muss sich bei diesem Vorgang automatisch an die ältere Version anpassen.	
Hub-Page	Unter dem Begriff Hub-Page versteht man eine WWW-Seite (einen hypertextualen Knotenpunkt), die semantisch kategorisierte Hyperlinks zu anderen Webseiten sammelt. Die Links sind thematisch und inhaltlich sortiert und in entsprechende Kategorien eingeordnet.	http://de.wikipedia.org/wiki/Link_hub
Hyperlink	Bei einem Hyperlink bzw. Link handelt es sich um ein Element zum Verweis bzw. zur Verknüpfung von Hypertext-Dokumenten. Hyperlinks werden vorwiegend zur strukturierten oder hierarchischen Vernetzung von Informationsstrukturen in Hypertexten verwendet. Befindet man sich auf einer Internetseite, so wird durch einen Maus-Klick auf einen Hyperlink bzw. Link eine weitere Internetseite in den Web-Browser geladen, die gegebenenfalls weiterführende Informationen zum Hyperlink-Begriff enthält oder mit diesem im Zusammenhang steht. Die gesamte Struktur des Internets basiert auf Hyperlinks bzw. Links. Ein Hyperlink bzw. Link erkennt man im Web-Browser daran, dass der Cursor sein Erscheinungsbild von einem Pfeil in eine Hand ändert. Die Funktion eines Hyperlinks besteht darin, die Navigation innerhalb von Hypertexten zu vereinfachen und auf einfache Weise auf weiterführende Informationen zu verweisen.	http://www.bullhost.de/h/hyperlink.html

Glossar I

	Definition	Quelle
If-modified-since	Durch das „If-Modified-Since“-Feld im HTTP-Protokoll ist technisch gesehen beim Übertragen einer WWW-Seite ein für die Aktualisierung wichtiger Mechanismus vorhanden. Über die Angabe dieses Feldes kann beim Laden eines Dokuments über das Internet die Übertragung von der letzten Änderung (Datum und Uhrzeit) abhängig gemacht werden. Das heisst, falls das Dokument seit dieser Zeitangabe geändert wurde, wird das Dokument übertragen, sonst nicht.	Tutorial Konstanz
IMAGE	IMAGE ist ein Datentyp, der von FTP unterstützt wird und auch als Übertragungsbefehl dient. IMAGE (oder BINARY) dient dabei zur Übertragung der Daten ohne Konvertierung, so dass die Dokumente auf Server und Client nach der Übertragung bitweise identisch vorliegen.	http://www.datenschutz-praxis.de/lexikon/ftp.html
IMAP	IMAP (Interactive Mail Access Protocol) dient dem Abruf und der Verwaltung von E-Mails auf dem Mail-Server. Hierbei werden die E-Mails auf dem Server in hierarchisch organisierten Postfächern verwaltet. IMAP wurde mit dem Ziel entworfen, den Zugriff auf Mailboxen und Nachrichten so bereitzustellen, als wenn diese sich auf dem lokalen Rechner befänden.	http://de.wikipedia.org/wiki/IMAP
Indexqualität	Die Indexqualität ist, neben der Trefferqualität, den	Lewandowski, D. &

	<p>Recherchemöglichkeiten und der Usability einer Suchmaschine, ein wichtiges Element für die Suchmaschinenevaluation. Die Indexqualität wird durch verschiedene Faktoren beeinflusst. Einen grossen Einfluss hat sicher der unterschiedlich gute Ranking-Algorithmus, doch auch der Aufbau und die Vollständigkeit (Abdeckung des Internet) und Aktualität des Index sind entscheidend.</p>	<p>Höchstötter, N. (2008): Web Searching. A Quality Measurement Perspective. In: Spink, Amanda & Zimmer, Michael (eds.). Web Search. Multidisciplinary Perspectives. Springer Series in Information Science and Knowledge Management 14. Berlin: Springer</p>
Information Retrieval	<p>Information Retrieval bzw. Informationswiedergewinnung, gelegentlich Informationsbeschaffung, ist ein Fachgebiet, das sich mit computergestütztem inhaltsorientiertem Suchen beschäftigt. Es ist ein Teilgebiet der Informationswissenschaft, der Computerlinguistik wie auch der Informatik. Wie der Begriff retrieval (deutsch Wiedergewinnung, Auffindung) sagt, sind Informationen in grossen Datenbeständen zunächst verloren und müssen wieder gewonnen bzw. wieder gefunden werden.</p>	<p>http://de.wikipedia.org/wiki/Information_Retrieval</p>
Intelligente Agenten	<p>Intelligente oder mobile Agenten sind Programme, die Personen oder Software als persönliche Assistenten dienen. Dazu benutzen sie Methoden der künstlichen Intelligenz und sind in der Lage, mit einer gewissen Autonomie komplexe Aufgaben zu übernehmen. Die Intelligenten Agenten sollen dem Benutzer helfen, mit dem „Information Overload“ effizient umzugehen und Zeit zu sparen. In der Regel besitzen Intelligente Agenten folgende Eigenschaften:</p> <ul style="list-style-type: none"> - Sie sind zielorientiert, der Benutzer kümmert sich um das „Was“ seiner Anfrage und nicht um das „Wie“. - Sie sind flexibel, je nach Umstand ändern sie selbständig ihr Verhalten (z.B. bei Veränderungen in der Softwareumgebung). - Sie „arbeiten mit“ und rechnen damit, dass der Benutzer auch Eingabefehler macht. - Sie können zu einem Selbststart fähig sein und z.B. ihre Suche dann starten, wenn die Netzbelastung gerade gering ist. - Sie sind kommunikationsfähig und können mit anderen Agenten oder anderer Software kommunizieren und so selbständig Informationen beschaffen. - Sie passen sich dem Benutzer an, kennen dessen Gewohnheiten und Arbeitsweisen. <p>Die Intelligenten Agenten werden sich wohl in naher Zukunft nicht manifestieren können. Es gibt verschiedene Kritikpunkte wie z.B. die Frage nach dem Datenschutz betreffend die Kommunikation mit anderen Agenten oder Programmen. Über allem steht die Frage nach dem Sinn der Intelligenten Agenten. Fraglich ist, inwieweit die Bedürfnisse der Benutzer wirklich erfasst und selbständig</p>	<p>http://de.wikipedia.org/wiki/Software-Agent</p> <p>http://www.stud.fernuni-hagen.de/q4559630/seminar1914/index.html</p>

	von den Agenten umgesetzt und auch vertreten werden können.	
Invisible Web	<p>Ein Grossteil der Dokumente des WWW ist für Suchmaschinen nicht erreichbar. Man spricht in diesem Zusammenhang von einem „Deep“ oder „Invisible Web“. Das „Invisible Web“ umfasst alle Inhalte, auf die aufgrund von Zugangsbeschränkungen durch die Anbieter oder technische Restriktionen seitens der Suchmaschinen nicht zugegriffen werden kann. Dies betrifft Dokumente, die z.B.:</p> <ul style="list-style-type: none"> - neu erstellt oder gerade aktualisiert wurden - nicht verlinkt sind und dadurch von keiner Suchmaschine gefunden werden können - abhängig sind von Benutzereingaben (diese können Roboter nicht ersetzen) - nur über eine Benutzererkennung erreichbar sind - oder durch den „Standard for Robot Exclusion“ bzw. Meta-Robots-Tag Zugangsgeschützt sind 	<p>Tutorial Konstanz</p> <p>http://de.wikipedia.org/wiki/Invisible_Web</p>

Glossar J

	Definition	Quelle
JPEG (Joint Photographic Experts Group)	Das JPEG-Format wurde von der Joint Photographic Experts Group, einem Gremium der ITU (International Telecommunication Union), entwickelt. Dieses Kompressionsverfahren ist verlustbehaftet, das Bild ist nicht mehr exakt rekonstruierbar, da einige Bildinformationen beim Speichern verloren gehen. JPEG-Dateien brauchen dadurch auch weniger Speicherplatz als andere Formate.	http://de.wikipedia.org/wiki/JPEG

Glossar K

	Definition	Quelle
Key Word Advertising	Siehe Key Word Sponsoring	
Key Word Sponsoring	Key Word Sponsoring (auch Key Word Advertising genannt) ist eine Werbeform bei Suchmaschinen. Dabei wird der Banner des Inserenten eingeblendet, sobald der zu bewerbende Suchbegriff vom Nutzer in das Suchfenster eingegeben wird.	http://de.wikipedia.org/wiki/Keyword-Advertising

Glossar L

	Definition	Quelle
Ländercode	Unter Ländercode (länderspezifische Top-Level-Domains) versteht man die Länderkennzeichnung für Internet-Adressen und Domain-Namen im Internet, wie z.B. „.ch“ für die Schweiz.	http://de.wikipedia.org/wiki/L%C3%A4ndercode
Link-Faktoren	Obwohl Suchmaschinen ihre Ranking-Methoden häufig geheim halten, kann man davon ausgehen, dass die Anzahl der Links, die auf eine Seite verweisen, einen entscheidenden Einfluss auf die Positionierung in der Trefferliste hat. Ein Link wird als Empfehlung betrachtet und je mehr solche Empfehlungen eine Seite hat, desto höher steigt sie im Ranking. Allerdings ist Link nicht gleich Link. Ein Eintrag im Webkatalog von Yahoo! ist beispielsweise mehr Wert als ein Link von	http://de.wikipedia.org/wiki/PageRank

	einer privaten Homepage. Google hat dazu ein eigenes Verfahren entwickelt, das PageRank.	
Link-Tag	Das HTML-<link>-Tag erzeugt eine Verbindung zwischen dem augenblicklich geöffneten Dokument und einem anderen, in irgendeiner Weise verwandtem Dokument. Das Link-Tag drückt somit Beziehungen zu anderen Quellen aus wie zum Beispiel zu einem Kapitel, Inhaltsverzeichnis, Stylesheet oder auch zu Angaben zum Ausgabemedium oder Zeichensatz.	http://de.selfhtml.org/html/kopfdaten/beziehungen.htm http://www.mediaevent.de/xhtml/link.html

Glossar M

	Definition	Quelle
Makroprecision	Makroprecision beschreibt, im Zusammenhang mit der Retrievaleffektivität der Suchmaschinen, wie effektiv die Suchmaschinen Informationsbedürfnisse befriedigen. Bei diesem Mass wird verglichen, welche Suchmaschine im Vergleich zur anderen mehr relevante Treffer liefert. Dementsprechend werden Rangplätze verteilt. Das Messen der Qualität der einzelnen Treffer bzw. der einzelnen Suchanfragen ist ein Problem und eine Lösung ist sehr schwierig zu finden, da alle Bewertungen eine gewisse Subjektivität besitzen.	Griesbaum, J. / Rittberger, M. / Bekavac, B. (2002): Deutsche Suchmaschinen im Vergleich. Alta.Vista.de, Fireball.de, Google.de und Lycos.de.
Mapping	Unter der EDV Bezeichnung Mapping versteht man das Zuordnen von Daten aus zwei verschiedenen Feldern, Speicherbereichen oder Protokollen zueinander. Typisches Beispiel ist das Adress-Mapping, bei dem die Programmadressen in Hardware-Adressen umgewandelt und im Hauptspeicher abgelegt werden, oder die Umwandlung von Adressen eines Protokolls in die eines anderen.	http://www.itwissen.info/definition/lexikon/mapping-Abbildung.html
Matching	Matching ist eine Suchmethode, bei der Begriffe, die das Informationsbedürfnis des Suchenden umschreiben, mit Begriffen, die aus den indextierten Dokumenten gewonnen werden, verglichen werden. Üblicherweise wird ein Matching anhand der Anzahl Treffer bewertet.	Tutorial Konstanz
MDA (Mail Delivery Agent)	Mit dem Mail Delivery Agent (MDA) verschickt der MTA (Mail Transfer Agent) E-Mails an eine bestimmte Mailbox des Benutzers. Jedes Programm, das eine Mitteilung zum Übermitteln an die Stelle, an der sie mit Hilfe eines MUAs (Mail User Agents) gelesen werden kann, verarbeitet, kann als MDA betrachtet werden. Viele Benutzer verwenden nicht direkt MDAs, da zum Versenden und Empfangen einer E-Mail nur MTAs und MUAs benötigt werden. Einige MDAs können jedoch zum Sortieren von Mitteilungen verwendet werden, bevor sie gelesen werden.	http://de.wikipedia.org/wiki/Mail_Delivery_Agent
Metasuchmaschine	Wichtig bei Metasuchdiensten ist, dass die Suchdienste nur über ein einziges Suchformular angesprochen werden. „Echte“ Metasuchdienste weisen folgende Charakteristiken auf: <ul style="list-style-type: none"> - mehrere Suchdienste, in der Regel meist Suchmaschinen und Kataloge, werden automatisch über eine Schnittstelle (Suchformular) befragt - die verschiedenen Suchdienste werden vorgegeben, können manchmal aber auch vom Benutzer ausgewählt werden 	Tutorial Konstanz

	<ul style="list-style-type: none"> - Funktionalität und Operatoren der verschiedenen Suchdienste werden verwendet. Hierbei wird eine Anpassung der Anfrage auf die einzelnen Suchdienste vorgenommen. Mindestens die Boole'schen Operatoren AND und OR sollten zur Verfügung stehen - die spezifischen Eigenschaften der unter der Meta-Maschine liegenden Suchdienste dürfen für die Bedienung keine Rolle spielen, der Anwender muss nichts darüber wissen - Kurzbeschreibungen der ausgewiesenen Suchtreffer z.B. als Titel oder Inhaltszusammenfassung werden übernommen und dargestellt - Eliminierung von Mehrfachtreffern aus den Ergebnissen der verschiedenen Suchdienste. Die Ergebnisse werden zusammengeführt und einheitlich dargestellt - Zeitvorgaben und maximale Treffergrenzen können gesetzt werden. Die Zeit für die Suche ist so einstellbar, dass der letzte nachgewiesene Treffer eines Suchdienstes noch erfasst werden kann. 	
Meta-Tags	<p>In der HTML-Sprache (hypertext markup language) besteht die Möglichkeit, Meta-Informationen in Meta-Tags getrennt vom Textteil („body“) im „head“ unterzubringen. Diese Informationen werden nicht am Bildschirm angezeigt, wenn ein solches Dokument durch einen Browser dargestellt wird; sie können jedoch durch den Befehl „view source“ oder mit einem Texteditor sichtbar gemacht werden. Heute werden Meta-Tags vor allem von den Suchmaschinen benutzt, die diese Tags auswerten können. Daneben gibt es spezielle Anweisungen, die den Suchroboter einer Suchmaschine steuern können. Einst galt das Optimieren der Meta-Tags als Geheimwaffe, um bei einer Suchmaschine möglichst weit oben gelistet/gerankt zu werden. Mittlerweile werten die Suchmaschinen wieder mehr den eigentlichen Textinhalt einer Internetseite aus, die auch der Betrachter dieser Seite auf einem Browser sieht.</p>	<p>http://dbs.schule.de/bonsai/metatags.html</p> <p>http://de.wikipedia.org/wiki/Meta-Tag</p>
Mikroprecision	<p>Bei der Mikromethode wird die Positionierung der einzelnen relevanten (inkl. verweisenden) Treffer untersucht. Damit wird eine Aussage über die Qualität des Ranking gewonnen.</p>	<p>Griesbaum, J. / Rittberger, M. / Bekavac, B. (2002): Deutsche Suchmaschinen im Vergleich. Alta.Vista.de, Fireball.de, Google.de und Lycos.de.</p>
MIME (Multipurpose Internet Mail Extensions)	<p>Über MIME (Multipurpose Internet Mail Extensions) kann der Inhalt einer Mail unterteilt (z.B. in attachments) und über verschiedene standardisierte Kennungen die zugehörigen Inhaltstypen beschrieben werden. Die ursprüngliche Herausforderung bestand darin, dass E-Mails nur mit dem ISO-7-bit Code (ASCII Zeichen von 0 bis 127) verschlüsselt wurden und damit eine Übertragung von binären Inhalten, wie Graphiken und Multimediadateien, zunächst nicht möglich war. Aus diesem Grund wurde eine Abbildung der für Binärdaten</p>	<p>Tutorial Konstanz</p> <p>http://de.wikipedia.org/wiki/Multipurpose_Internet_Mail_Extensions</p>

	<p>notwendigen 8-bit Kodierung auf die 7-bit Kodierung der E-Mails erforderlich. MIME wird hierfür als Standard genutzt und ermöglicht den Einbezug von Binärdaten in den Body der E-Mail, indem zusätzliche Konfigurationsfelder eingefügt werden. Diese Binärdaten geben unter anderem Auskunft über den Typ der nachfolgenden Daten. Dadurch wird der Body der eigentlichen Email in mehrere Teile segmentiert. Es existieren sowohl eine Anzahl standardisierter MIME Content Types, wie text/plain (ASCII-Text), text/html (HTML-Text), application/zip (mit dem ZIP-Packer komprimierte Dateien), image/jpeg (JPEG-Graphik) und multipart/mixed (aus mehreren unabhängigen Datentypen zusammengesetzter Bodypart), als auch Möglichkeiten, individuelle Datentypen, gekennzeichnet mit einer mit „X-“ beginnenden Attributbezeichnung, zu definieren.</p>	
Mirroring	<p>Das Mirroring oder Spiegeln von Festplatten ist dort sinnvoll, wo ein Datenbestand permanent zu jeder Zeit zur Verfügung stehen soll (z.B. Daten eines WWW-Servers). Bei Ausfall einer Festplatte gehen die Daten nicht verloren, da sie ja auf anderen Festplatten redundant gespiegelt sind und damit eine Sicherheitskopie existiert.</p>	<p>http://de.wikipedia.org/wiki/Mirroring</p>
Mobile Agenten	Siehe intelligente Agenten	
MTA (Mail Transfer Agent)	<p>Ein Mail Transfer Agent (MTA) ist ein Programm, das den Transport und die Verteilung von E-Mail erledigt. Für die Nutzung von E-Mail benötigt man ein lokales Mailprogramm, welches in der Fachwelt als Message User Agent (MUA) bezeichnet wird, sowie mindestens einen Mailserver zum Transport der E-Mail vom Sender zum Empfänger. Ein solcher Mailserver wird als Mail Transfer Agent (MTA) bezeichnet. Solche Server sind in mehrere Komponenten unterteilbar. Eine Komponente dient der Weiterleitung der E-Mail über das Internet und eine Komponente ist zuständig für die Zustellung der E-Mail an die entsprechende Email-Adresse des oder der Empfänger. Die letztere Komponente wird auch als Mail Delivery Agent (MDA) bezeichnet. Welcher Mail Transfer Agent für bestimmte E-Mails zuständig ist, wird aufgrund des Domain-Adressbestandteils der den Mails zugeordnet ist, geregelt. Das Kommunikationsprotokoll für den Austausch von E-Mails über das Internet ist das Simple Mail Transfer Protocol (SMTP).</p>	<p>Tutorial Konstanz</p> <p>http://de.wikipedia.org/wiki/Mail_Transfer_Agent</p>
MUA (Message User Agent)	<p>Für die Nutzung von E-Mail benötigt man ein lokales Mailprogramm, welches in der Fachwelt als Message User Agent (MUA) bezeichnet wird, sowie mindestens einen Mailserver zum Transport der E-Mail vom Sender zum Empfänger. Ein solcher Mailserver wird als Mail Transfer Agent (MTA) bezeichnet. Der MUA dient zur Bearbeitung von E-Mails durch den Benutzer. Das Kommunikationsprotokoll für den Austausch von E-Mails über das Internet ist das Simple Mail Transfer Protocol (SMTP).</p>	<p>Tutorial Konstanz</p> <p>http://de.wikipedia.org/wiki/Mail_Transfer_Agent</p>
MX-Record (Mail Exchanger Resource Record)	<p>Die Spezifikation des DNS (Domain Name Servers) definiert für besondere Zwecke des Mailroutings sogenannte Mail Exchanger Resource Records (MX-Records). Es handelt sich hierbei um spezielle Resource Records für das Mailrouting. Ein MX-Record spezifiziert einen Domainnamen und einen zugehörigen Rechner, den sogenannten Mail Exchanger, der in der Lage ist,</p>	<p>http://www.mathematik.uni-ulm.de/~ffischer/IFCvortrag/page3.html</p>

	Mail an eine E-Mail-Adresse mit diesem Domainnamen auszuliefern. Es ist möglich, mehrere Mail Exchanger für einen Domainnamen anzugeben. Mailsysteme können so konfiguriert werden, dass sie einen Nameserver nach Mail-Exchanger-Einträgen für den Domain-Teil einer gegebenen Zieladresse abfragen.	
--	---	--

Glossar N

	Definition	Quelle
NewsFeeds	NewsFeeds sind NNTP (Network News Transport Protocol)-Server. Ein NNTP-Server kann News-Artikel sowohl von anderen Servern oder den als Newsreader bezeichneten Clients der Nutzer empfangen, als auch an weitere Server weitersenden. Artikel, die bereits versandt wurden, werden nicht weiter berücksichtigt. Dadurch soll eine rasche Verbreitung der geposteten Artikel gewährleistet werden. Eine weitere Funktion besteht in der Bereitstellung gewünschter News-Artikel für die News-Clients der Nutzer. Die Artikel sind getrennt nach Header für die Adressierung und Body mit dem eigentlichen Artikel.	Tutorial Konstanz
NewsGroups	Bei NewsGroups handelt es sich um ein System von öffentlichen schwarzen Brettern, die nach Themen bzw. Foren unterteilt sind. Die NewsGroups dienen dem Austausch von Nachrichten/Artikeln. Dies geschieht häufig über das NNTP-Protokoll (Network News Transport Protocol). In ihrer Gesamtheit können die NewsGroups fast jedes denkbare Thema abdecken, wobei vor allem aktuelle Ereignisse von Bedeutung sind. Um die riesige Menge an Themen und Foren übersichtlich zu strukturieren, sind diese in hierarchische Kategorien aufgeteilt. Weltweit existieren seit Jahrzehnten verschiedene News-Systeme. Das bekannteste ist dabei das Usenet, das unzählige Newsserver weltweit verbindet. Oft sind Newsgroups aber auch rein regional verteilt, oder werden in Intranets zur Verfügung gestellt. Newsgroups sind wie E-Mail ein Dienst im Internet. Das Usenet ist deutlich älter als das WWW. Zum Lesen von Newsgroups benötigt man einen Newsreader. Ausserdem können sie über Web-Interfaces wie bspw. Google Groups (http://groups.google.de) gelesen werden.	Tutorial Konstanz http://de.wikipedia.org/wiki/Newsgroup
Newsreader	Ein Newsreader (auch Newsclient) ist ein Computerprogramm zum Lesen und Schreiben (und je nach Bedarf auch wieder löschen) von Nachrichten in Newsgroups des Usenets. Die Newsreader-Programme müssen meist zuerst auf dem Rechner installiert werden. Neben echten Newsreadern gibt es auch Schnittstellen, die Newsgroups im WWW abbilden und daher auch dessen Nutzung mit Browsern erlauben. Ein Beispiel ist http://groups.google.com/ . Die Nutzung dieser Services wird von vielen Lesern nicht gern gesehen, da hier zum einen bestehende Standards nicht beachtet werden und zum anderen diese Dienste des öfteren von Spammern missbraucht werden.	http://de.wikipedia.org/wiki/Newsreader
NNTP	NNTP steht für Network News Transfer Protocol. Das	http://de.wikipedia.org/

(Network News Transport Protocol)	Protokoll regelt den Austausch von Usenet- Artikeln in Computer-Netzwerken. NNTP definiert eine Anzahl von Befehlen für die interaktive Übertragung und den Download von News-Artikeln. NNTP ist ein Client-Serverorientiertes Protokoll.	org/wiki/NNTP Tutorial Konstanz
--	---	---

Glossar O

	Definition	Quelle
OCR (Optical Character Recognition)	Bei OCR (Optical Character Recognition - optische Zeichenerkennung) handelt es sich um ein Verfahren zur Erkennung von Text, der mittels eines Scanners eingelesen wurde. Die beim Scannen erzeugte Bitmap- oder Pixel-Grafik wird dabei mittels einer speziellen OCR-Software in Textinformationen umgewandelt.	http://www.bullhost.de/o/ocr.html
On-Page-Faktoren	Die On-Page-Faktoren beeinflussen, neben weiteren Aspekten wie den On-Site-Faktoren und den Link-Faktoren, die Position eines Dokuments im Ranking einer Suchmaschine. Dabei werden insbesondere die folgenden Aspekte des Dokuments analysiert: <ul style="list-style-type: none"> - Häufigkeit, Position (Dichte, Abstand) - Funktion (URL, HTML-Auszeichnungen: Titel, Überschriften, Linktexten...) - Format von Termen (Schriftgröße, Farbe) Die Relevanz eines Dokuments wird z.B. erhöht, je häufiger Anfrageterme in einem Dokument vorkommen. Diese Faktoren sind für das Ranking jedoch nicht ausreichend. Zum einen suchen die Nutzer meistens mit nur einem Suchbegriff und zum anderen können diese Faktoren vom Websitebetreiber manipuliert werden. Deshalb werden bei der Bewertung z.B. auch die Link-Faktoren einer Website ausgewertet.	Griesbaum, J. / Bekavac, B. / Rittberger, M. (2008): Typologie der Suchdienste im Internet. In: Dirk Lewandowski (Hrsg.), Handbuch Internet-Suchmaschinen - Nutzerorientierung in Wissenschaft und Praxis. Heidelberg: Akademische Verlagsgesellschaft.
On-Site-Faktoren	Die On-Site-Faktoren dienen, wie die On-Site-Faktoren, der Platzierung eines Dokuments im Ranking einer Suchmaschine. Dabei werden globale Faktoren der jeweiligen Domain, auf der sich die Dokumente befinden, analysiert. Die Art der verwendeten On-Site-Faktoren und ihre reale Bedeutung sind aber weitgehend unbekannt. Denkbar ist u.a., dass neben dem Alter der Domain auch Faktoren wie die Art der Domain, ihre Linkpopularität, die thematische Ausrichtung der Gesamtsite, die Gesamtzahl der indexierten Seiten usw. bereits jetzt oder künftig herangezogen werden.	Griesbaum, J. / Bekavac, B. / Rittberger, M. (2008): Typologie der Suchdienste im Internet. In: Dirk Lewandowski (Hrsg.), Handbuch Internet-Suchmaschinen - Nutzerorientierung in Wissenschaft und Praxis. Heidelberg: Akademische Verlagsgesellschaft.
Open Directory Project (ODP)	Das Open Directory Project (ODP), auch bekannt als DMoz (für „Directory at Mozilla“), ist das umfangreichste von Menschen erstellte Open Content Link-Verzeichnis des WWW. Es wird bearbeitet und aktualisiert von einer Gemeinschaft freiwilliger Editoren. Die Daten des ODP werden als Open Content zu den Bedingungen der Open Directory Project-Lizenzbedingungen (http://dmoz.org/license.html) zur Verfügung gestellt, welche die Aufnahme einer ODP-Quellenangabe (http://dmoz.org/become_an_editor/) mit Links zur	http://de.wikipedia.org/wiki/Dmoz

	<p>Siteanmeldung und Editorbewerbung auf jeder einzelnen Verzeichnis-Seite vorschreibt. Die Daten stehen öffentlich zum Download zur Verfügung (http://rdf.dmoz.org/). Zu den Nutzern der ODP-Daten gehören neben Netscape und AOL auch bekannte Suchmaschinen wie Google oder Excite sowie mehr als 400 weitere Sites. Die ursprüngliche Motivation zur Gründung von ODP war die Frustration über die mangelhafte Aktualität, das heisst, die lange Verzögerung bei der Aufnahme neuer Sites und der hohe Anteil „toter Links“, in redaktionell betreuten Verzeichnissen.</p> <p>Im Juni 2008 verzeichnete das ODP über 4,5 Millionen Einträge. Seit der Gründung des Projektes haben mehr als 80.000 Editoren zum Wachstum des Verzeichnisses beigetragen.</p>	
OPAC (Online Public Access Catalog)	Ein OPAC (Online Public Access Catalog) ist ein öffentlich zugänglicher digitaler Bibliothekskatalog.	http://de.wikipedia.org/wiki/Bibliothekskatalog
Open Access	<p>Open Access ist ein wissenschaftspolitisch inspiriertes Ziel, die wissenschaftliche Literatur im Internet frei zugänglich zu machen. Dazu gibt es eine internationale Bewegung. Der Schwerpunkt liegt auf der Forderung, dass alle wissenschaftlichen Arbeiten (Artikel wie Monografien und alle anderen Formen) kostenlos zugänglich sein sollen. Die entstehenden Kosten werden durch unterschiedliche Geschäftsmodelle von den Produzenten getragen. Open Access hat neben den finanziellen aber auch technische, inhaltliche und rechtliche Bedingungen.</p> <p>Open Access-Zeitschriften sind ausschliesslich wissenschaftliche Publikationen, die ebenso einer Qualitätsprüfung unterliegen (können) wie traditionelle gedruckte Zeitschriften, die in einem Verlag publiziert werden. Darüber hinaus bieten Open Access Zeitschriften Autoren verbesserte Möglichkeiten bezüglich ihrer Rechte beim Publikationsprozess, da sie ihre Rechte beim Veröffentlichen nicht übertragen müssen.</p>	http://de.wikipedia.org/wiki/Open_Access http://open-access.net/de/allgemeines/was_bedeutet_open_access/

Glossar P

	Definition	Quelle
PageRank	<p>Beim PageRank-Verfahren von Google wird der Hyperlink-Struktur der vernetzten Web-Dokumente eine zentrale Rolle zugestanden. Dabei wird vor allem die Anzahl eingehender Verweise (Backlinks) auf ein bestimmtes Dokument betrachtet: Je häufiger dieses von anderen Web-Seiten referenziert wird, desto wahrscheinlicher ist es, dass es sich dabei um ein inhaltlich hochwertiges Dokument handelt. Dieses Verfahren der Linkpopularität wird praktisch von allen Suchmaschinenanbietern verwendet. Das PageRank-Verfahren geht jedoch noch einen Schritt weiter. Bei diesem Verfahren spielt auch die „Qualität“ der Referenzen selbst eine wichtige Rolle. Sind diese Dokumente bereits schon als „hochwertig“ eingestuft worden oder befindet sich der Link in bekannten Web-Katalogen bzw. auf einschlägigen Webseiten, so wird dies ebenfalls zu Gunsten des jeweiligen Dokuments</p>	<p>Bekavac, B. (2004). Metainformationsdienste des Internet. In: Kuhlen, R. und Seeger, T. und Strauch, D. (Hrsg.), Grundlagen der praktischen Information und Dokumentation. Saur Verlag.</p> <p>Griesbaum, J. & Bekavac, B. (2004a). Von der Kommerzialisierung bis zum Deep Web:</p>

	<p>ausgelegt. Es wird sogar der Text der Hyperlinks bzw. der Text im Umfeld von Hyperlinks auf Begriffsebene analysiert, um weitere Anhaltspunkte für die Relevanz der Dokumente zu erhalten.</p> <p>Der grosse Erfolg bzw. Bekanntheitsgrad von Google basiert insbesondere auch auf dem Page-Rank Verfahren. Durch dieses Verfahren kann eine sehr hohe Relevanz (Precision) im vorderen Bereich der Trefferliste erzielt werden. Ein weiterer Vorteil ist, dass eine Manipulation, wie sie durch Spam-Techniken bei einem Ranking aufgrund von statistischen Kriterien möglich ist, hierbei einen sehr hohen Aufwand erfordert respektive kaum möglich ist. Nachteilig ist, dass neue Dokumente mit wenigen Verweisen bei diesem Verfahren prinzipiell benachteiligt werden und somit eine Tendenz zur Verstetigung der Suchergebnisse besteht.</p>	<p>Problemfelder der Internetsuche. Universität Konstanz / HTW Chur.</p>
Paid Inclusion	<p>Mit Paid Inclusion ist die gegen Barzahlung garantierte Aufnahme und regelmässige Aktualisierung im Index einer Suchmaschine gemeint. Die über Paid Inclusion erfassten Seiten unterliegen dem normalen Rankingverfahren der Suchmaschine, eine verbesserte Position kann also nicht erkaufte werden. Es können allerdings mehrere Seiten einer Homepage aufgenommen werden, wodurch sich indirekt die Rankingposition verbessern kann.</p> <p>Anhand dieses Verfahrens können auch Inhalte bereit gestellt werden, die unter Umständen von den Suchmaschinen nicht gefunden werden - also auch Inhalte des Deep Web.</p>	<p>Tutorial Konstanz</p>
Paid Placement / Paid Listing	<p>Paid Placement oder Paid Listing ist das Versteigern von Suchbegriffen. Wer für einen bestimmten Suchbegriff den höchsten Betrag zu zahlen bereit ist, der wird im Ranking an erster Stelle geführt. Wer den zweithöchsten Betrag bietet kommt auf Rang zwei etc. Es handelt sich also um eine Positionsersteigerung. Die Bezahlung erfolgt durch das so genannte „pay-per-click“-Verfahren. Das heisst, der Anbieter einer Seite muss den von ihm gebotenen Preis erst bezahlen, wenn seine Seite von einem Benutzer der Suchmaschine angeklickt wird. Ein Beispiel für eine solche Suchmaschine war Overture. Die Adresse www.overture.com verlinkt neu auf eine Yahoo-Site, auf welcher für die „Sponsored Search“ geworben wird. Das Paid Listing wird vor allem im Bereich B-to-B angewendet. Firmen können so gezielt ihr Produkt vermarkten und Geschäftsbeziehungen pflegen. Ausserdem können auf diese Weise „Deep-Links“ gesetzt werden, die direkt auf die Produkte oder Dienstleistungen einer Firma zugreifen.</p>	<p>Tutorial Konstanz</p> <p>http://www.trafficm. axx.de/sponsored-links.htm</p>
Paid Submission	<p>Mittlerweile ist es vor allem bei Webverzeichnissen und -katalogen üblich, eine Gebühr für die Bearbeitung der Anmeldung zu erheben. Dieses System bezeichnet man als „Paid Submission“. Je nach Zahlungswilligkeit des Anbieters kann dadurch auch die Bearbeitungszeit der Anmeldung verkürzt und eventuell die Frequenz der späteren Aktualisierung erhöht werden. Im Gegensatz zur Positionsersteigerung wird dem Anbieter einer Webseite aber keine bestimmte Position im Ranking zugesichert. Der Betreiber behält sich zudem vor, die Seite trotz Bezahlung abzulehnen. Unter welcher Rubrik</p>	<p>Tutorial Konstanz</p>

	ein Katalogeintrag erfolgt, entscheidet ebenfalls die Redaktion. Beim Webkatalog Yahoo wurde in der Vergangenheit die Gebühr bereits schon allein für die Begutachtung einer Homepage fällig („Pay for Consideraton“).	
Pay for Consideration	Pay for Consideration heisst, dass bei der Anmeldung bei Webverzeichnissen und Webkatalogen eine Gebühr allein schon für die Begutachtung einer Homepage fällig wird.	Tutorial Konstanz
Pay per Click	Der Begriff Pay per Click taucht in der Positionsersteigerung von Suchbegriffen auf (Siehe Paid Placement, Paid Listing). Pay per Click heisst für den Positionsersteigerer, dass die Gebühr an den Suchdienstbetreiber erst dann fällig wird, wenn der Link des Positionsersteigerers tatsächlich angeklickt wird - egal, ob dadurch ein Geschäft zustande kommt oder nicht.	Tutorial Konstanz
PDF (Portable Document Format)	Das Adobe Portable Document Format (PDF) ist der offene De facto-Standard für die elektronische Dokumentenverteilung weltweit. PDF ist ein universelles Dateiformat, das alle Schriften, Formatierungen, Farben und Grafiken eines beliebigen Quelldokuments beibehält. Dies geschieht unabhängig von dem Programm und dem Betriebssystem, mit dem es erstellt wurde. PDF-Dateien sind kompakt und können von allen Nutzern im festgelegten Layout betrachtet, und exakt ausgedruckt werden. Dies erfordert den kostenlosen Adobe Acrobat Reader. Es können beliebige Dokumente in das PDF-Format konvertiert werden.	http://de.wikipedia.org/wiki/Portable_Document_Format
Phrasensuche	Mit der Phrasensuche bezeichnet man die Suche nach der exakten Reihenfolge der angegebenen Suchbegriffe. Um eine Phrasensuche einzuleiten, müssen die Suchbegriffe meist in Hochkomma eingeschlossen sein. Es existieren dazu jedoch auch oft vordefinierte Suchfelder mit Bezeichnungen wie „den genauen Ausdruck“ suchen.	Tutorial Konstanz
Platzhalter	Siehe Wildcard	
POP3 (Post Office Protocol Version3)	Das Post Office Protocol (POP) ist ein Protokoll für den Nachrichtenaustausch durch E-Mail. Es gestattet das Abrufen der auf einem zentralen Mailserver in den jeweiligen Email-Accounts eingeordneten Nachrichten durch den lokalen Message User Agent (MUA). Beim Zugriff werden die Mails lokal heruntergeladen und auf dem Server gelöscht. Um dabei einem Missbrauch vorzubeugen, ist für das Herunterladen von E-Mails auf den eigenen Rechner eine Authentifizierung mittels Nutzererkennung und Passwort notwendig.	Tutorial Konstanz
Portal	Portale bilden eine spezielle Ausprägung von Web-Katalogen bzw. Verzeichnissen und gelten als Einstiegspunkte in das WWW. Portale werden meistens von bekannten Providern oder Suchdiensteanbietern, wie z.B. Yahoo! bereitgestellt. Ein Portal kann anhand folgender Kriterien von „einfachen“ Katalogen unterschieden werden: <ul style="list-style-type: none"> - Portale sollten betrachtet werden als Einstiegspunkt für den vereinfachten Zugang zu einer grossen Menge an Informationen und Angeboten des durch sie repräsentierten Mediums, z.B. Internet, Intranet. 	Tutorial Konstanz

	<ul style="list-style-type: none"> - Portale verwenden solche Informationsquellen und Suchwerkzeuge, die den wirtschaftlichen Interessen der Portal-Anbieter entsprechen. Dazu zählen unterschiedliche Dienste wie Suchmaschinen, Kataloge oder Spezialsuchdienste für bestimmte Themen oder Datentypen. Zusätzlich ist auch eigener, redaktionell aufgearbeiteter Inhalt („Content“) zu bestimmten Themengebieten aus eigenen oder fremden Quellen charakteristisch für Portale. - Ein wichtiges Merkmal von Portalen sind Möglichkeiten der Personalisierung zur Anpassung der Seiten des Portals an persönliche Interessen und Präferenzen. Typisch wären auch Ergänzungen als Tools zur Unterstützung von netzbasierten Aktivitäten und Kommunikation z.B. Email-Account, Bookmarkmanagementtool, Adressbuch, Kalender, themenspezifische Chatrooms usw. <p>Portale können ferner unterteilt werden in horizontale Portale, die inhaltlich allgemein gefasst sind, und vertikale Portale, die auf bestimmte Themen oder Zielgruppen fokussieren.</p>	
Postscript	<p>Postscript ist eine Seitenbeschreibungssprache, die seit 1982 von der Firma Adobe entwickelt wird. Grafiken und Druckseiten werden als Dateien im PostScript-Format angelegt, um sie auf den unterschiedlichsten Ausgabegeräten, in variabler Grösse verlustfrei auszugeben. Graphische Elemente und Schriften werden in PostScript mit Hilfe von Vektoren beschrieben. Pixel-Graphiken werden gerastert. Postscript ist optimiert für die Druckerausgabe und eignet sich nur bedingt für die Anzeige am Bildschirm. Postscript-Dateien sind ausserdem nicht für die Bearbeitung gedacht, sondern ein endgültiges Ausgabeformat.</p> <p>Postscriptfähige Drucker und Druckmaschinen sind zu diesem Zwecke mit einem PostScript-Interpreter ausgestattet, der die Datei zeilenweise interpretiert. Man unterscheidet Hard- und Software-Interpreter. Eine freie Software-Implementierung eines solchen Interpreters bietet die Software Ghostscript. Postscript hat sich über die Jahre zu einem Standard in der Druckindustrie entwickelt. Das weit verbreitete PDF baut auf PostScript auf, verwendet jedoch nur einen eingeschränkten Befehlssatz. Es fehlen insbesondere die Strukturen einer Programmiersprache.</p>	http://de.wikipedia.org/wiki/PostScript
Precision	<p>Die Precision beschreibt mit dem Anteil relevanter Dokumente an der Ergebnismenge die Genauigkeit eines Suchergebnisses.</p> <p>Precision ist zusammen mit Recall ein Mass zur Beurteilung der Güte von Treffermengen einer Recherche beim Information Retrieval. Beide Masse können Werte zwischen Null und Eins (beziehungsweise 0% bis 100%) annehmen und hängen voneinander ab. Idealerweise sollten zum einen alle Treffer relevant (Precision: hoch) sein, zum anderen sollten alle relevanten Webseiten in der Trefferliste enthalten (Recall: hoch) sein. Dieser Idealfall ist in der Realität kaum zu erreichen: Die Erhöhung des einen Werts hat die Verringerung des anderen zur Folge.</p>	http://de.wikipedia.org/wiki/Recall_und_Precision

	Für die Bewertung eines Rechercheverfahrens werden beide Werte gemeinsam betrachtet, z. B. in einem Precision-Recall-Diagramm. Die Berechnung ist allerdings meist nicht problemlos möglich, da nicht alle Werte bekannt sind: Während die Precision sich nach der Auswertung einer überschaubaren Treffermenge leicht errechnen lässt, bleibt Recall ein theoretisches Konstrukt, da sich die für ihre Berechnung erforderlichen Werte in der Praxis nicht ermitteln lassen.	
Provider	Provider geht auf das lateinische Verb providere (provideo, providi, provisum) zurück und bedeutet im Deutschen soviel wie „Lieferer“ oder „Versorger“. In der Telekommunikation wird mit Provider ein Anbieter von Telekommunikationsdiensten bezeichnet. Das können höherwertige Dienste sein, aber auch einfache Dienste zum Internet. Besitzt der Provider ein eigenes Kommunikationsnetz, wird er als Netzbetreiber bezeichnet. Internetdiensteanbieter (engl. Internet Service Provider) werden im deutschsprachigen Raum auch oft als Provider bezeichnet. Dieser bietet in der Regel gegen Entgelt verschiedene technische Leistungen an, die für die Nutzung oder den Betrieb von Internet-Diensten erforderlich sind.	http://de.wikipedia.org/wiki/Provider http://de.wikipedia.org/wiki/Internetdiensteanbieter
Proximity Operators	Proximity Operators (Abstandsoperatoren) werden bei der Eingabe von Suchbegriffen verwendet. Anhand von Proximity Operators können der Umkreis, in dem die Begriffe liegen sollen, oder auch die exakte Reihenfolge der Begriffe vorgegeben werden.	Tutorial Konstanz
Proxy-Server	Ein Proxy oder Proxyserver (vom engl. proxy representative = Stellvertreter von lat. „proximus“ = „Der Nächst(e)“) ist ein Computerprogramm, das im Datenverkehr zwischen Computern oder Computer-Programmen in so genannten Computernetzen zwischen angefragtem Server und anfragendem Client vermittelt. Dem Server gegenüber verhält sich das Programm wie ein Client, dem Client gegenüber wie ein Server. Strukturell/logisch liegt der Proxyserver zwischen anfragendem Client und dem angefragten Server, zu dem er vermittelt. Der strukturell Nächste ist hier aber nicht notwendigerweise auch der räumlich Nächste.	http://de.wikipedia.org/wiki/Proxyserver

Glossar Q

	Definition	Quelle
Quoted-printable	Quoted-Printable ist ein Verfahren, mit dem man in E-Mails Zeichen, die nicht im ASCII-Zeichensatz enthalten sind, kodieren kann. Dies ist erforderlich, da bis heute nicht sichergestellt ist, dass alle Zwischenstellen beim E-Mail-Versand sogenannte 8-Bit-Zeichen verarbeiten können; die ursprüngliche E-Mail-Spezifikation sah nur 7 Bit vor. Um trotzdem deutsche Umlaute oder andere Sonderzeichen wie das Euro-Symbol verwenden zu können, ohne dabei die Kompatibilität zu älteren Systemen zu gefährden, werden diese Zeichen kodiert. Dafür gibt es verschiedene Verfahren, die im MIME-Standard spezifiziert sind. Die Quoted-Printable-Kodierung hat den Vorteil, dass die meisten in ASCII	http://de.wikipedia.org/wiki/Quoted_printable

	enthaltenen Zeichen unverändert kodiert werden und so auch angezeigt werden können, wenn die Quoted-Printable-Kodierung vom Empfänger nicht unterstützt wird.	
--	---	--

Glossar R

	Definition	Quelle
Ranking	<p>Unter dem englischen Terminus Ranking (deutsch: Einstufung) versteht man die Darstellung von Ergebnissen in Form von Ranglisten. Rankings ermöglichen es, Informationsangebote nach bestimmten Kriterien zu bewerten. Auch bei der Abfrage von Suchmaschinen werden die Ergebnisse nach einem Ranking dargestellt, nämlich in einer Reihenfolge, wo die ersten Treffer die höchste Relevanz für den Suchbegriff bilden. Eines der Kriterien für die Einstufung ist die Zahl der Verweise anderer Webseiten auf eine bestimmte Homepage. Je mehr Links im Internet auf die betreffende Webseite verweisen, desto höher wird sie eingestuft, weil eine höhere Übereinstimmung zwischen Suchbegriff und dem Inhalt der Website angenommen wird.</p> <p>Welche Kriterien für das Ranking der Suchmaschinen herangezogen werden, ist meist ein Firmengeheimnis der jeweiligen Suchmaschine (z.B. PageRank von Google).</p>	<p>http://de.wikipedia.org/wiki/Suchmaschinenranking</p> <p>http://de.wikipedia.org/wiki/Ranking</p>
Rankingalgorithmen	<p>Der Rankingalgorithmus ist eine mathematische Formel zur Berechnung der Ranking-Position einer Webseite in Bezug auf eine bestimmte Suchanfrage. Eine Formel zur Relevanzbestimmung eines Webdokuments in Relation zu der an eine Suchmaschine gestellten Frage.</p> <p>Jede Suchmaschine verwendet eigene Algorithmen, mit dem Ergebnis, dass die gleiche Webseite bei verschiedenen Suchmaschinen sehr unterschiedlich gelistet sein kann. Von Google ist bekannt, dass über 100 Faktoren in die Berechnung des Rankings einfließen.</p>	<p>http://www.sumo.de/definition-ranking-algorithmus.html</p>
Recall	<p>Der Recall gibt mit dem Anteil der bei einer Suche gefundenen relevanten Dokumente die Vollständigkeit eines Suchergebnisses an.</p> <p>Recall ist zusammen mit Precision ein Mass zur Beurteilung der Güte von Treffermengen einer Recherche beim Information Retrieval. Beide Masse können Werte zwischen Null und Eins (beziehungsweise 0% bis 100%) annehmen und hängen voneinander ab. Idealerweise sollten zum einen alle Treffer relevant (Precision: hoch) sein, zum anderen sollten alle relevanten Webseiten in der Trefferliste enthalten (Recall: hoch) sein. Dieser Idealfall ist in der Realität kaum zu erreichen: Die Erhöhung des einen Werts hat die Verringerung des anderen zur Folge.</p> <p>Für die Bewertung eines Rechercheverfahrens werden beide Werte gemeinsam betrachtet, z.B. in einem Precision-Recall-Diagramm. Die Berechnung ist allerdings meist nicht problemlos möglich, da nicht alle Werte bekannt sind: Während die Precision sich nach der Auswertung einer überschaubaren Treffermenge leicht errechnen lässt, bleibt Recall ein theoretisches Konstrukt, da sich die für ihre Berechnung</p>	<p>http://de.wikipedia.org/wiki/Recall_und_Precision</p>

	erforderlichen Werte in der Praxis nicht ermitteln lassen.	
received:	„received:“ ist ein Feld im E-Mail-Header. In diesem Feld kann der zurückgelegte Weg einer E-Mail anhand der dort protokollierten Einträge von einzelnen Servern nachvollzogen werden.	Tutorial Konstanz
Relevanz	<p>In der Informationswissenschaft ist die Frage der Relevanz von Dokumenten ein wichtiger Punkt bei der Informationswiedergewinnung. Dabei geht es darum, in welchem Umfang die gefundenen Informationen dem Informationsbedürfnis entsprechen.</p> <p>Es gilt zu unterscheiden zwischen subjektiver Relevanz (Pertinenz) und objektiver Relevanz.</p> <p>Ein Dokument ist für eine Suchanfrage (objektiv) relevant,</p> <ul style="list-style-type: none"> - wenn es objektiv zur Vorbereitung einer Entscheidung dient. - wenn es objektiv eine Wissenslücke schliesst. - wenn es objektiv eine Frühwarnfunktion erfüllt. <p>Suchmaschinen nutzen Kriterien der Relevanz, um Dokumente bei der Ausgabe zu sortieren ("Relevance Ranking"). Dabei spielen für den Benutzer ebenfalls die Aktualität, die Quelle und der Neuigkeitswert der Informationen eine grosse Rolle.</p>	<p>http://de.wikipedia.org/wiki/Relevanz_(Informationswissenschaft)</p> <p>http://en.wikipedia.org/wiki/Relevance_(information_retrieval)</p>
Retrievaleffektivität	Siehe Makroprecision & Mikroprecision	
Robots-Exclusion-Standard	<p>Der Robots-Exclusion-Standard bietet die Möglichkeit, WWW-Server vor Roboterzugriffen zu schützen. Dies funktioniert über eine spezielle Datei im Serververzeichnis, über die Roboter erfahren, welche Dokumente bzw. Teile des Servers gelesen werden dürfen. Der Robots-Exclusion-Standard ist jedoch kein echter Zugriffsschutz, sondern eine Art Abmachung unter den Betreibern der Suchdienste.</p> <p>Es gibt zwei Möglichkeiten um den Zugriff für Roboter auf Server zu sperren. Eine erste Möglichkeit ist die Datei „robots.txt“, welche bestimmte Server-Bereiche die Durchsuchung von Robotern sperrt. Eine zweite Möglichkeit ist das META-Robots-TAG im HTML-Header, welches dem Roboter mitteilt, welche Seiten er indexieren darf und welche Seiten nicht. Diese Sperrmechanismen funktionieren jedoch nur, sofern sich der Roboter an das Protokoll hält.</p>	<p>Tutorial Konstanz</p> <p>http://de.wikipedia.org/wiki/Robots_Exclusion_Standard</p>
RSS	<p>RSS steht für „Rich Site Summary“ oder „RDF Site Summary“ oder „Really Simple Syndication“. Es handelt sich um einen XML-Standard, um Textinformationen für andere Rechner zur Verfügung zu stellen. In einer RSS-Datei wird die Information auf die Überschrift und einen Link reduziert. Diese werden logisch gekennzeichnet. Zur Nachricht kommt der Leser, indem er dem Link folgt. Der Einsatz eines RSS-Readers spart Zeit, da man mehrere News- oder Weblog-Seiten abonnieren kann und die zugehörige News- oder Weblog-Seiten abrufen. Spam gibt es nicht mehr, in Abonnentenlisten taucht man nicht mehr auf.</p>	http://de.wikipedia.org/wiki/RSS

Glossar S

	Definition	Quelle
Sampling	Sampling ist die Digitalisierung eines analogen Vorgangs. Ein bekanntes Beispiel ist das Aufnehmen von Klängen in computerlesbarer Form. Töne etc. werden zunächst analog aufgezeichnet, das heisst vom Mikrofon in Spannungsschwankungen umgesetzt. Diese Signale werden beim Sampling digitalisiert, z.B. durch den ADC-Baustein einer Soundkarte. Dabei werden die Schwingungskurven in regelmässigen Zeitabständen abgetastet. Für bestimmte Punkte dieser Kurve ergeben sich jeweils Werte (Bits), die zugeordnet und gespeichert werden. Da sehr viele Punkte berücksichtigt werden, entsteht ein Strom von Daten (Bitstrom). Die Qualität dieser digitalisierten Aufzeichnung hängt von der Samplingrate und der Samplingtiefe ab.	http://www.kublikon.de/lexikon/html/s.php
Schnittstelle	Eine Schnittstelle (englisch interface) ist ein Teil eines Systems, das dem Austausch von Informationen, Energie oder Materie mit anderen Systemen dient. Eine Schnittstelle wird durch eine Menge von Regeln beschrieben, der Schnittstellenbeschreibung. Neben der Beschreibung von Funktionen und deren Benutzung, gehört zur Schnittstellenbeschreibung auch ein so genannter Kontrakt. Im Kontrakt wird die Semantik der einzelnen Funktionen beschrieben.	http://de.wikipedia.org/wiki/Schnittstelle
SemanticWeb	Das SemanticWeb ist eine Erweiterung des World Wide Web. Das Konzept beruht auf einem Vorschlag von Tim Berners-Lee, dem Begründer des World Wide Webs. Ziel des SemanticWebs ist es, die Bedeutung von Informationen für Computer verwertbar zu machen. Durch das SemanticWeb sollen Informationen durch Maschinen automatisch miteinander in Beziehung gesetzt werden können. Bei einer Reise etwa Wetterdaten und Staumeldungen mit Informationen über Orte und Abfahrtszeiten von Zügen. Das SemanticWeb setzt Annotationen (Metadaten) voraus. Die Bedeutung der dargebotenen Inhalte wird also mit Hilfe einer Auszeichnungssprache explizit dazugeschrieben und nicht erst später heuristisch interpretiert wie in der Computerlinguistik.	http://de.wikipedia.org/wiki/Semantisches_Web
Serendipity-Effekt	Wer browsst ist zunächst einmal ganz unbefangen auf Entdeckungsreise in der „globalen Bibliothek“ des World Wide Web. Diese Entdeckungsreise kann mehr oder weniger zielgerichtet sein. Selbst wenn ein User sein Browsing unter der anfänglichen Motivation beginnt, eine Ressource zu einem ganz bestimmten Thema finden zu wollen, tritt diese Motivation oftmals in den Hintergrund, wenn er beim Herumstöbern auf eine Ressource zu einem ganz anderen Thema stösst. Dies bezeichnet man als „Serendipity-Effekt“. Der Ausdruck „Serendipity“ leitet sich her von Serendip, dem alten Namen der Insel Ceylon (des heutigen Sri Lanka) und ist motiviert durch eine Erzählung von Horace Walpole mit dem Titel „The three kings of Serendip“, in welchem es um drei Könige geht, die beständig Dinge finden, die sie gar nicht gesucht haben.	http://www.hypermedia-texte.de/hyper2/navi.html
Sitemap	Der Begriff Sitemap hat im Zusammenhang mit dem Internet zwei unterschiedliche Bedeutungen. Einerseits wird als Sitemap die vollständige hierarchisch	http://de.wikipedia.org/wiki/Sitemap

	<p>strukturierte Darstellung aller Einzeldokumente (Webseiten) eines Internetauftritts bzw. einer Website bezeichnet (Seitenstruktur). Andererseits steht Sitemap für Metadaten-Informationen eines Webauftritts für Suchmaschinen. Dabei handelt es sich um eine einfache Textdatei mit einer Auflistung der Seitenadressen oder eine XML-Datei, welche die URLs der Webseiten einschliesslich einiger Metadaten (letzte Aktualisierung, Änderungshäufigkeit, Relevanz, Relation zu anderen URLs des Webauftritts) enthält, so dass die Suchmaschinen den Webauftritt leichter durchsuchen können.</p> <p>Im Jahre 2006 haben Google, Yahoo! und Microsoft gemeinsam beschlossen, einen Standard festzulegen, wie Sitemap-Informationen strukturiert sein müssen, damit Suchmaschinen diese Informationen maschinell auslesen können. Ziel ist eine Verbesserung der Suchresultate.</p>	
SMTP (Simple Mail Transfer Protocol)	<p>Die Abkürzung SMTP steht für Simple Mail Transfer Protocol und ist ein Protokoll der TCP/IP-Protokollfamilie, das den Versand von E-Mails in Computer-Netzwerken regelt. SMTP ist also ein Kommunikationsprotokoll für den Austausch von E-Mails über das Internet. Ein MTA (Mail Transfer Agent) kann Mails damit sowohl empfangen als auch selbst weitersenden. Der zurückgelegte Weg einer E-Mail kann anhand der im Header unter der Kennung „received:“ protokollierten Einträge der einzelnen Server nachvollzogen werden. SMTP setzt voraus, dass eine Übertragung vom Sender initiiert wird, aus diesem Grund wird es nicht dazu benutzt, eine Mail von einem Server auf den Arbeitsplatzrechner zu übertragen. Dazu werden Post Office Protokolle wie das POP3-Protokoll, das IMAP-Protokoll oder andere verwendet.</p>	<p>http://de.wikipedia.org/wiki/SMTP</p> <p>Tutorial Konstanz</p>
Social Bookmarkdienste	<p>Zu den Informationsanbietern der neueren Generation gehören Social Bookmarkdienste oder sogenannte Tagging Communities. Sie sind nicht eigentliche Informationslieferanten sondern erreichen durch die Beteiligung ihrer User eine Ansammlung von Informationsquellen zu bestimmten Themen. Durch die inhaltliche und semantische Erschliessung werden so Quellen aus Literatur und Web zu einem Webkatalog zusammengefügt und Drittpersonen bzw. anderen Users zur Verfügung gestellt.</p> <p>Wichtige Dienste dazu sind Flickr (Bildercommunity) und Delicious (Social Bookmarks). Solche Dienste werden als ein zentraler Entwicklungstrend des Web2.0 gesehen.</p>	<p>Griesbaum, J. / Bekavac, B. / Rittberger, M. (2008): Typologie der Suchdienste im Internet. In: Dirk Lewandowski (Hrsg.), Handbuch Internet-Suchmaschinen - Nutzerorientierung in Wissenschaft und Praxis. Heidelberg: Akademische Verlagsgesellschaft.</p>
Spam (E-Mail)	<p>Spam ist die gängige Bezeichnung für unverlangt zugesandte Werbepost per E-Mail (Unsolicited commercial E-Mail; kurz: UCE).</p>	<p>http://de.wikipedia.org/wiki/Spam</p>
Spam (Suchmaschine)	<p>Spam ist der Missbrauch der Suchmaschinen-Spider, um eine Seite möglichst weit oben im Ranking einer Suchmaschine zu platzieren. Es ist nicht einheitlich definiert, was als Spam gilt. Jeder Suchdienst entscheidet selber, welche Methoden zugelassen sind und welche nicht. Es gibt verschiedene Arten von Spam:</p> <ul style="list-style-type: none"> - Wiederholung von Keywords, Elementen (Metatags, Bilder) 	<p>Tutorial Konstanz</p> <p>http://de.wikipedia.org/wiki/Suchmaschinen-Spamming</p>

	<ul style="list-style-type: none"> - Webseiten mit vielen Wörtern, die nur für die Platzierung im Ranking sind - Themenfremde Keywords - Versteckter Text (gleiche Farbe wie Hintergrund oder in Bildern) - Wiederholungen vom Seitentext - Weiterleitungen - Mehrfaches Anmelden einer Seite beim selben Dienst - Cloaking - Linkfarmen <p>Spam ist ein grosses Problem für Suchmaschinen, da ihre Spider „ausgetrickst“ werden und es nicht ohne weiteres möglich ist, dies zu verhindern.</p>	
Spider	<p>Spider, auch Roboter, Wanderer oder Worms genannt, sind Programme, welche die Webseiten indexieren und auf Aktualisierungen prüfen. Sie „wandern“ über Hypertext-Verlinkungen von Seite zu Seite. Dafür gibt es zwei Prinzipien:</p> <ul style="list-style-type: none"> - „Depth first“: Der Spider verfolgt einen Link über mehrere Seiten hinweg in die „Tiefe“. - „Breadth first“: Es werden zuerst alle abgehenden Links von einer Seite verfolgt. <p>Mit Hilfe von Spidern können auch statistische Analysen von Webservern oder Dokumenten erstellt werden. Sie können Dead-Links aufspüren, gespiegelte Seiten und verschiedene Dokumentformate erkennen und finden. Für die Aktualisierung der indexierten Seiten gibt es die sogenannte Spiderfrequenz, also die Zeit, die der Spider benötigt, um wieder auf dieselbe Seite zu gelangen. Die Spiderfrequenz sollte möglichst kurz sein, was jedoch selten der Fall ist. Meist dauert es mehrere Monate. Wenn viele Spider mit hoher Frequenz auf einen Server zugreifen, kann ausserdem die Serverleistung beeinträchtigt werden.</p>	<p>Tutorial Konstanz</p> <p>http://de.wikipedia.org/wiki/Webcrawler</p>
Spider-Frequenz	Siehe Spider	
Spiegeln	Siehe Mirroring	
Stoppwörter	<p>Stoppwörter nennt man Wörter, die bei einer Volltextindexierung nicht beachtet werden, da sie sehr häufig auftreten und gewöhnlich keine Relevanz für die Erfassung des Dokumentinhalts besitzen. Allgemein übliche Stoppwörter in deutschsprachigen Dokumenten sind bestimmte Artikel (der, die, das, unbestimmte Artikel (einer, eine, ein), Konjunktionen (z.B. und, oder, doch) und häufig gebrauchte Präpositionen (z.B. an, in, von), sowie die Negation „nicht“. Abhängig von den zu erschliessenden Dokumenten können Stoppwörter auch mehrsprachig vorliegen. Diesen Wörtern ist gemeinsam, dass sie vor allem grammatische/syntaktische Funktionen übernehmen und daher keine Rückschlüsse auf den Inhalt des Dokumentes zulassen. Eine weitere Gemeinsamkeit ist ihre grosse Zahl: Sie treten in jedem Dokument sehr zahlreich auf und kommen in sehr vielen Dokumenten vor.</p> <p>Das Weglassen von Stoppwörtern dient der Steigerung der Effizienz von Suchmaschinen. Würde man Stoppwörter bei einem Suchauftrag beachten, würde die Ergebnismenge nahezu jedes Dokument des Bestandes enthalten. Das Suchergebnis wäre unscharf, sein Nutzwert gering.</p>	<p>http://de.wikipedia.org/wiki/Stoppwort</p>

Suchmaschinenmarketing (SEM)	Suchmaschinenmarketing (Search Engine Marketing, SEM) umfasst alle Massnahmen zur Gewinnung von Besuchern für eine Website über Suchmaschinen. Dazu gehören vor allem Optimierungsmassnahmen, welche das Ranking einer Website verbessern. Es gibt aber auch die Möglichkeit des Key Word Advertising. Das heisst für bestimmte Suchbegriffe wird eine Position auf der ersten Seite der Suchmaschinenergebnisse gekauft.	http://de.wikipedia.org/wiki/Suchmaschinenmarketing
Suchmaschinenoptimierung (SEO)	Search Engine Optimization (SEO) (engl. für Suchmaschinenoptimierung) ist ein Fachbegriff für Massnahmen, die dazu dienen, dass Webseiten auf den Ergebnisseiten von Suchmaschinen auf höheren Plätzen erscheinen. Damit das geschieht vergibt man z.B. Metatags (Anweisungen für die Suchmaschinen-Roboter) und achtet bei der Gestaltung der Seite z.B. auf Farben, Grösse etc. (= On-Page-Optimierung). Daneben spielt die Platzierung einer Seite (wo wird sie angemeldet) sowie natürlich die Verlinkungen von und zur Seite eine wichtige Rolle (= Off-Page-Optimierung).	http://de.wikipedia.org/wiki/Suchmaschinenoptimierung
Suchraumeingrenzung	Durch die Suchraumeingrenzung bei einer Suchmaschinen-Anfrage kann die Treffermenge eingeschränkt bzw. die Menge an relevanten Treffern erhöht werden. Eingrenzungen sind nach verschiedenen Kriterien möglich: <ul style="list-style-type: none"> - Sprache - Land - Dokumentart - Dateiformat - Bei Bildern: nach Bildgrösse - Vorkommen der Suchbegriffe in bestimmten Feldern - Zeitraum - Nutzungsrechte/Lizenz - Informationen für Erwachsene/ Jugendfreie Informationen - Quellen - Fachgebiete oder Rubriken - Klang der Suchbegriffe 	http://www.google.ch/advanced_search?hl=de
Surfen	Als Surfen, oder auch Internetsurfen, wird umgangssprachlich das Ansehen von mehreren Webseiten in Folge bezeichnet. Man kann hier zwischen zwei Varianten unterscheiden: Einerseits das wahllose Aufrufen von Webseiten, bei dem in der Regel der Inhalt der besuchten Seiten keine Rolle spielt, und andererseits das Folgen von Links von Webseite zu Webseite, wobei in der Regel nach bestimmten Informationen gesucht wird oder Interesse an bestimmten Themengebieten besteht.	http://de.wikipedia.org/wiki/Internetsurfen

Glossar T

	Definition	Quelle
Tagging Communities	Siehe Social Bookmarkdienste	
Tags	Tags sind Markierungselemente zum Beispiel in HTML und XML. Dabei werden Textelemente in paarweise (öffnende und schliessende) Tags eingeschlossen. Die jeweils zusammengehörenden Tags bilden ein Element. Diese Elemente lassen sich nach Regeln, die in einer Dokumenttypdefinition (DTD) angegeben sind,	http://de.wikipedia.org/wiki/HTML

	<p>verschachteln. Neben Elementen mit Start- und End-Tag gibt es auch leere Elemente, wie etwa Zeilenumbrüche oder Bilder.</p> <p>Dabei sind diese Tags keine Präsentations-Befehle (obwohl das in Zeiten von HTML 3.2 so war), die dem interpretierenden Webbrowser mitteilen, er müsse beispielsweise bis zum Auftreten des Endtags alles in Fettschrift setzen, sondern eine strukturgebende Auszeichnung, deren Repräsentation von der Umgebung abhängig ist. Obwohl HTML-Dokumente in der Regel auf Computerbildschirmen dargestellt werden, kann man sie auch auf anderen Medien ausgeben lassen, etwa auf Papier oder als Sprachausgabe.</p>	
Telnet	<p>Telnet ist der Name eines im Internet weit verbreiteten Protokolls. Der Sinn des TELNET-Protokolls besteht darin eine ziemlich allgemeine, bi-direktionale, 8-Bit-pro-Byteorientierte Kommunikationsmöglichkeit zu bieten. Es wird üblicherweise dazu verwendet, Benutzern den Zugang zu Internetrechnern über die Kommandozeile zu bieten. Aufgrund der fehlenden Verschlüsselung wird es mehr und mehr durch andere Protokolle verdrängt. Weiterhin ist Telnet der Name eines Programms, welches Verbindungen zu einem entfernten Gastrechner ermöglicht. Das Telnetprogramm stellt dabei die benötigten Clientfunktionen des Protokolls zur Verfügung. Telnet ist ein Client-Server-Protokoll. Es verwendet TCP und die Clients verbinden sich meistens über Port 23 mit dem Zielrechner. Allerdings lässt sich dieser Port wie bei den meisten Internetprotokollen auch ändern. Teilweise aufgrund des Protokolldesigns, teilweise aufgrund der üblicherweise zur Verfügung gestellten Flexibilität der Telnetprogramme, ist es ebenfalls möglich, mit einem Telnetprogramm eine interaktive TCP-Verbindung zu einigen anderen Internetservices aufzubauen. Eine klassische Verwendung davon ist beispielsweise eine Telnetverbindung über Port 25 aufzubauen (wo sich meistens ein SMTP-Server befindet) um Fehler in einem EMail-Server zu finden.</p>	<p>http://de.wikipedia.org/wiki/Telnet</p>
Thumbnail	<p>Mit Thumbnail (engl. Daumennagel) wird eine Mini-Grafik bezeichnet, die oft als Vorschau-Bild für eine grössere Grafik oder ein grösseres Bild dient. Erst durch anklicken des Thumbnails wird dann das grössere Bild geladen. Thumbnails sind gut geeignet für Übersichten von Bildern und Bildgalerien und werden auch für die Ergebnisdarstellungen von Bildersuchen in einigen Suchmaschinen dargestellt.</p>	<p>http://www.symweb.de/glossar/thumbnail_379.htm</p>
Top10 Mean Average Precision	<p>Bei der Evaluation von Suchmaschinen ist unter anderem das Ranking ein wichtiger Faktor. Die Methode der Top10 Mean Average Precision konzentriert sich auf die Qualität des Rankings. Der entsprechende Wert wird aus dem Anteil relevanter Treffer an der Gesamtmenge von Treffern für jede Trefferposition von 1-10 berechnet. Dabei wird der Durchschnitt sämtlicher Suchanfragen errechnet.</p>	<p>Griesbaum, J. / Rittberger, M. / Bekavac, B. (2002): Deutsche Suchmaschinen im Vergleich. Alta.Vista.de, Fireball.de, Google.de und Lycos.de.</p>
Trunkierung	<p>Mit Trunkierung oder Wortstammsuche bezeichnet man die Suche nach verschiedenen Wortvariationen. Die Trunkierung erlaubt das Ersetzen eines oder mehrerer</p>	<p>Tutorial Konstanz</p>

	<p>Buchstaben zu Beginn oder am Ende eines Wortes bspw. durch das *-Zeichen.</p> <p>Beispiel:</p> <ul style="list-style-type: none"> - hand* - sucht nach allen Worten, die mit „hand“ anfangen: Hand, Handschuh, Handlung... - *hand - sucht nach allen Worten, die mit „hand“ enden: Treuhand... 	
--	--	--

Glossar U

	Definition	Quelle
UNIX	<p>UNIX ist heute der Oberbegriff für alle Betriebssysteme, die auf dem ursprünglichen, von AT&T (American Telephone & Telegraph Corporation) entwickelten Unix-Entwurf basieren. Zu diesen Unix-Derivaten gehört zum Beispiel Mac OS X. Der Einfluss von Unix erstreckt sich auf nahezu alle heute verbreiteten Betriebssysteme. Ideen und Konzepte von Unix finden sich überall. Sei es im kleinen Embedded System oder im Grossrechner-Betriebssystemen. Projekte wie GNU/Linux oder die freien BSD (Berkeley Software Distribution)-Systeme implementieren das Verhalten und die Schnittstellen von Unix neu und steuern eigene Erweiterungen bei. Das System wurde ursprünglich ab 1969 an den Bell Laboratories von AT&T von einer Gruppe um Ken Thompson und Dennis Ritchie entwickelt und in den Jahren 1972-1974 bis auf wenige Teile in der höheren Programmiersprache C mit dem Ziel der Portabilität neu implementiert. UNIX ist ein Open-Source-Betriebssystem.</p>	<p>http://de.wikipedia.org/wiki/Unix</p>
URL (Unified Resource Locator)	<p>Einfach formuliert ist der Uniform Resource Locator (URL engl. „einheitlicher Ortsangeber für Ressourcen“) der Name eines Webservers im Internet, über den dieser angesprochen werden kann. Der URL wird zusammen mit der IP-Adresse dieses Rechners im DNS (Domain Name System) registriert. Ein URL ist ein Uniform Resource Identifier (URI), der eine Ressource über ihren primären Zugriffsmechanismus, das heisst dem Ort (engl. location) der Ressource im Internet, identifiziert. Der Begriff URL wird (historisch bedingt) oft synonym zum Begriff Uniform Resource Identifier (URI) verwendet, obwohl es sich bei URLs nur um eine Unterart von URIs handelt.</p>	<p>http://de.wikipedia.org/wiki/Uniform_Resource_Locator</p>
Usability (Suchmaschine)	<p>Grundsätzlich haben Suchdienste die allgemeinen Usability-Anforderungen an Webseiten zu erfüllen. Da ein Suchdienst aber eine aktive Interaktion mit dem Nutzer eingeht, müssen hier weitere Punkte beachtet werden:</p> <ul style="list-style-type: none"> - Die Informationen über den Zweck und Umfang der Suchmaschine sollten transparent sein. - Der durchschnittliche Nutzer sollte nicht gezwungen werden irgendwelche Operatoren zu benutzen. Das System sollte ihm diese Arbeit abnehmen. - Die Suchmaschine sollte nachsichtig sein und allfällige Tippfehler erkennen und korrigieren. - Bei zu vielen oder keinen Treffern sollte die Suchmaschine sinnvolle Hilfestellung bieten wie 	<p>Schulz, U. (2007). [Web Usability] Kriterien für Suchmaschinen. URL: http://www.bui.haw-hamburg.de/pers/ursula.schulz/webusability/suchma.html</p>

	<p>z.B. Filter.</p> <ul style="list-style-type: none"> - Bei der Trefferliste ist es wichtig, dass diese schnell aufgebaut wird und die Treffermenge angibt. 	
--	---	--

Glossar V

	Definition	Quelle
Virus	<p>In der Fachsprache ist ein Computervirus eine nichtselbständige Programmroutine, die sich selbst reproduziert, indem sie sich an andere Computerprogramme oder Bereiche des Betriebssystems anhängt. Ist der Virus einmal gestartet, kann er vom Anwender nicht kontrollierbare Manipulationen am Computer vornehmen. Umgangssprachlich hat der Begriff Computervirus eine breitere Bedeutung: Er wird sowohl für Computerviren in der fachsprachlichen Bedeutung als auch für Computerwürmer und Trojanische Pferde benutzt. Die Idee zu Computerviren leitete sich von dem biologischen Vorbild der Viren ab. Durch Computerviren kommt es auf einem Computer häufig zur Veränderung oder Verlust von Daten und Programmen sowie zu Störungen des regulären Betriebs.</p>	http://de.wikipedia.org/wiki/Computervirus
Visible Web	<p>Unter dem Begriff Visible Web (deutsch: „Sichtbares Web“) versteht man denjenigen Teil des World Wide Web, der allgemein zugänglich ist. Es besteht also aus denjenigen Webseiten und Dokumenten, die von konventionellen Suchmaschinen, welche mit Hilfe von Crawlern über weiterführende Hyperlinks von einer Webseite zu weiteren URLs gelangen, erfasst werden können. So können theoretisch alle erreichbaren Seiten des WWW gefunden werden. Fakt ist jedoch, dass das Visible Web schätzungsweise 400 bis 550 mal kleiner als das Deep Web ist, was bedeutet, dass konventionelle Suchmaschinen lediglich die Spitze eines riesigen und immer weiter wachsenden Eisberges durchsuchen und somit bei der Recherche sichtbar machen.</p>	http://de.wikipedia.org/wiki/Surface_Web
Visualisierungen (Suchmaschinen)	<p>Visualisierungen dienen dazu, komplexe Sachverhalte und Datenstrukturen in einer Form darzustellen, die es dem Anwender ermöglicht, inhärente Zusammenhänge besser zu verstehen. Bei Suchmaschinen werden die folgenden Visualisierungsansätze verfolgt:</p> <ul style="list-style-type: none"> - Ergebnisvorschau: Die Ergebnisse werden zusätzlich mit einem kleinen Screenshot/Thumbnail dargestellt. Bsp: www.Exalead.com - Ergebnismengenvergleich: Die Rankings von mehreren Webseiten werden miteinander verglichen. Bsp: www.langreiter.com/exec/yahoo-vs-google.html - Clustering: Die Ergebnisse werden in grobe Kategorien unterteilt. Bsp: www.grokker.com - Relationennachweis: Die Vernetzungen (Hyperlinksverknüpfungen) der verschiedenen Ergebnisse werden dargestellt. Bsp: www.kartoo.com - Visualisierung mit kollaborativen Elementen: 	<p>Weinhold, T. / Bekavac, B. / Hierl, S. / Öttl, S. (2008): Visualisierung bei Internetsuchdiensten. In: Dirk Lewandowski (Hrsg.), Handbuch Internet-Suchmaschinen - Nutzerorientierung in Wissenschaft und Praxis. Heidelberg: Akademische Verlagsgesellschaft.</p>

	<p>Der Suchmaschinenbenutzer kann die Ergebnisse bewerten, taggen oder kommentieren. Bsp: www.eurekster.com</p> <ul style="list-style-type: none"> - Unterstützung der Suchanfragenformulierung: Der Suchmaschinenbenutzer wird durch Begriffsvorschläge unterstützt. Bsp: www.quintura.com - Visualisierung zeitliche Aspekte: Die Ergebnisse werden auf einer Zeitleiste dargestellt. Bsp: www.viewzi.com (Google Timeline) - Visualisierung geographische Aspekte: Die Ergebnisse werden zusätzlich auf einer Karte angezeigt. Bsp: http://maps.google.de/ 	
Visualisierungspipeline	Die Visualisierungspipeline spezifiziert die Prozesskette mittels derer Daten in Bilder überführt werden. Sie besteht aus in Reihe geschalteten Funktionen zum Generieren, Filtern und Bereinigen von Daten, zum Abbilden der Daten auf Geometrien und Materialien, zum Rendern dieser Objekte und zum Darstellen des gerenderten Bildes.	http://de.wikipedia.org/wiki/Visualisierung#Visualisierungspipeline

Glossar W

	Definition	Quelle
Wanderer	Siehe Spider	
Web 2.0	Der Begriff „Web 2.0“ bezieht sich weniger auf spezifische Technologien oder Innovationen, sondern primär auf eine veränderte Nutzung und Wahrnehmung des Internets: Die Benutzer erstellen und bearbeiten Inhalte in quantitativ und qualitativ entscheidendem Masse selbst. Die Inhalte werden nicht mehr nur zentralisiert von grossen Medienunternehmen erstellt und über das Internet verbreitet, sondern auch von einer Vielzahl von Nutzern, die sich mit Hilfe sozialer Software zusätzlich untereinander vernetzen. Typische Beispiele hierfür sind Wikis, Blogs, Foto- und Videoportale (z.B. Flickr und YouTube), soziale Online-Netzwerke wie Facebook, sowie Social-Bookmarking-Portale wie Delicious.	http://de.wikipedia.org/wiki/Web_2.0
Weblog	Ein Weblog (auch Blog genannt) ist eine Wortkombination aus „Web“ und „Logbuch“. Für „Weblog“ existieren keine allgemein gültigen Definitionen. Technisch gesehen sind Weblogs Seiten, die regelmässig aktualisiert werden und ihre Inhalte datiert und in umgekehrt chronologischer Reihenfolge auflisten (die aktuellsten Einträge stehen zu Beginn der Seite). Allgemein formuliert lässt sich sagen, dass in Weblogs einer oder mehrere Autoren (sogenannte „Blogger“) die Stationen ihrer Surftouren durch das Internet schriftlich kommentiert festhalten. Typisch ist auch die Weiterverlinkung auf die besuchten Seiten. Bezüglich Inhalten und Autorschaft bestehen jedoch grosse Unterschiede. Das Spektrum reicht von Autoren, die in ihren Weblogs persönliches festhalten, d.h. ein öffentliches Tagebuch führen, bis hin zu Fach-Weblogs, in denen die Autoren nur Artikel zu einem bestimmten Thema verfassen. Viele Weblog-Einträge bestehen aus Einträgen anderer	http://de.wikipedia.org/wiki/Weblog

	Weblogs oder beziehen sich auf diese. Somit besteht eine starke Vernetzung der Weblogs untereinander. Die Gesamtheit aller Weblogs wird „Blogosphäre“ genannt. Auch betreffend die Technik von Weblogs gehen die Meinungen auseinander. Es stellt sich die Frage, ob Weblogs mit fertig eingerichteter Software oder individueller Technik betrieben werden sollen. Es gibt eine Vielzahl an Diensten, bei denen man sich Weblog-Accounts anlegen kann.	
Wildcard	Wildcards oder Platzhalter sind Zeichen (meist „*“), die als Stellvertreter für einen Buchstaben oder eine Buchstabengruppe eingesetzt werden können. Wildcards eignen sich z.B. wenn man sich bei einer Schreibweise nicht ganz sicher ist. Bsp.: „m*ller“ - sucht nach allen Worten, die mit „m“ anfangen und mit „ller“ enden: Müller, Muller, Möller...	Tutorial Konstanz http://www.atweb.de/grundlagen/platzhalter.htm
Worm (Schadprogramme, Viren)	Ein Wurm oder Computerwurm ist eine selbständige Programmroutine, die sich selbst reproduziert, indem sie über ein Computernetzwerk an Computerprogrammen oder Betriebssystemen anderer Computern Manipulationen vornimmt. Ein Wurm kann eine spezielle Schadensroutine enthalten, muss aber nicht. Da ein Wurmprogramm auf befallenen Systemen Ressourcen zur Weiterverbreitung bindet, können selbst Würmer ohne spezielle Schadensroutinen gewaltige wirtschaftliche Schäden erzeugen. Heutzutage verbreiten sich Würmer überwiegend per E-Mail, wobei sie sich als Datei-Anhang an einen kurzen Text anfügen. Verschiedene Mechanismen dienen zum Tarnen des gefährlichen Anhangs bzw. zum automatischen Ausführen. Würmer bekommen Dateinamen mit doppelten Erweiterungen, z.B. music.mp3.exe.	http://de.wikipedia.org/wiki/Computerwurm
Worm (Roboterbasierte Suche)	Siehe Spider	
Wortstammsuche	Siehe Trunkierung	

Glossar X

	Definition	Quelle
XML (Extensible Markup Language)	XML (Extensible Markup Language) ist eine Metasprache für die Definition von Dokumenttypen. XML ist eine vereinfachte Variante von SGML (SGMLlight), zugeschnitten auf das Internet. Durch XML sollen vor allem mehr Metainformationen in die Dokumentstruktur integriert werden. Dies ermöglicht eine einfachere maschinelle Weiterverarbeitung. HTML wurde durch XML redefiniert (HTML 4.01 => XHTML 1.0). Dokumente auf der Basis von XML können auf mehreren DTDs (Document Type Definitions) basieren. Die Standardisierung der XML-DTDs übernimmt das W3C.	http://de.wikipedia.org/wiki/XML
XSL (Extensible Stylesheet Language)	XSL (Extensible Stylesheet Language) ist eine Familie von Sprachen zur Erzeugung von Layouts für XML-Dokumente. Diese Layouts (auch Stylesheets genannt) können in die zu formatierenden XML-Dokumente eingebunden werden, wobei sich die Layouts speziellen	http://de.wikipedia.org/wiki/XSL

	Medien zuordnen lassen. So ist es möglich, ein Layout zum Drucken und ein Layout für die Darstellung am Computer zu verwenden.	
--	--	--