

Finden von Informationen im Netz v1.3

Mattes Sarcander <mtsrc@systemausfall.org>

Zusammenfassung

Das menschliche Gedächtnis hat nur ein begrenztes Fassungsvermögen. Ungeachtet dieser Tatsache wächst die Menge an Daten, die vielen Menschen zur Verfügung steht, exponentiell. Für ein effizientes Arbeiten wäre es also sinnvoll, aktuelles Wissen weniger auf Vorrat zu lernen, sondern vielmehr die vorhandenen Werkzeuge zur bedarfsorientierten Recherche zu nutzen. Eine Einführung in die dafür benötigten Techniken soll dieser Text bieten.

Inhaltsverzeichnis

1	Einleitung	3
1.1	changelog	3
2	Grundlagen der Suche	4
2.1	Allgemeine Tipps	5
3	Kataloge	5
3.1	Yahoo	6
3.2	The Virtual Library	6
3.3	dmoz open directory project	7
4	Suchmaschinen	7
4.1	Google	7
4.1.1	Googles Filter	8
4.2	AlltheWeb	9
4.3	Paperball	9
5	Usenet	10
5.1	Google Groups	10
6	Hilfsmittel	10
6.1	Mailinglisten	10
6.2	Web Archive	11
A	Anhang	11
A.1	Suchen für Fortgeschrittene	11
A.2	Software	11
A.2.1	Browser: Firefox	11
A.2.2	Mailtool + Newsreader: Sylpheed	11
A.3	Glossar	12
A.4	Links	13
A.5	Lizenz	14

1 Einleitung

Durch die elektronischen Medien hat sich unsere Wissensverwaltung fundamental geändert. Früher gab es nur zentrale Anlaufstellen, Bibliotheken, die von Menschen verwaltet wurden. Der Mensch auf der Suche nach Erkenntnis ging zu den BibliothekarInnen und äußerte den Wunsch. Die BibliothekarInnen kannten das Archiv entweder auswendig, oder sie hatten wenigstens Kenntnis darüber, wo ein Buch zu dem Thema liegen könnte. Es gab also jemanden, der gefragt werden konnte, der auch dann den Sinn der Frage verstand, wenn die Fachbegriffe fehlten, und sichere Weisung zu den Dokumenten gab.

Dann kamen Computer.¹

Anfangs spielten sie als Archive praktisch keine Rolle. Als das World Wide Web gegen 1990 offiziell zugänglich wurde, kam vielleicht alle paar Tage eine neue Seite dazu, langsam genug, um sie durchzulesen und zu sehen, ob sie nützliche Informationen enthält. Doch bald wuchs das Web immer schneller, und das manuelle Durchstöbern wurde unpraktisch. Von nun an gab es zwei Entwicklungen:

Einerseits wurde versucht, die in der Netzwelt nicht vorhandenen BibliothekarInnen dadurch zu ersetzen, dass die neuen Seiten per Hand in Kategorien eingeteilt wurden. Daraus sind die Kataloge entstanden.

Die andere Option war es, den Computer suchen zu lassen. Dazu lesen Programme die Webseiten, und schreiben die Inhalte in Datenbanken. Die Nutzenden können nun bestimmte Begriffe in einem Formular eingeben, und erhalten als Antwort die Dokumente, die diese Begriffe enthalten. Diese Variante wird Suchmaschine genannt. Die Suchmaschinen "verstehen" die Anfrage nicht, dafür können die Archive aber bedeutend größer werden (Maschinen sind eben schneller).

Es kommt also darauf an, die Suchanfrage so zu gestalten, dass ein Programm sie bearbeiten kann. Es hat keinen Sinn, zu hoffen, dass das Programm bestimmte Andeutungen versteht oder Schlüsse zieht. Die Einführung dazu, wie solche Anfragen bei den unterschiedlichen Suchmaschinen und Katalogen zu formulieren sind, soll dieser Text bieten.

Es folgen jetzt ein kleiner Ausflug in die zugrundeliegende Theorie, dann einige allgemeine Hinweise, und im Hauptteil schließlich wird auf die Besonderheiten der einzelnen Suchmaschinen eingegangen. Unbekannte Begriffe können im Glossar im Anhang nachgelesen werden. Zur deutlichen Unterscheidung von anderen Inhalten werden auf den nachfolgenden Seiten jene Daten, die Ein- oder Ausgaben des Rechners darstellen, mit 'Hochkommata' geklammert, es ist empfehlenswert, sie bei einer Eingabe zu ignorieren.

1.1 changelog

- v1.3: 2005-2-25
neues zu Googles Filter
- v1.2: 2004-5-17
viele Tippfehler beseitigt und Formatierung optimiert
neue Sylpheed- und DFN-Adressen.
- v1.1: 2004-5-15
Googles Filter
empfohlenen Browser von Phoenix auf Firefox geändert
Lizenz angegeben
- v1: 2003-9-9
erster Release.

¹A Brief History of the Internet
<http://www.isoc.org/internet/history/brief.shtml>

2 Grundlagen der Suche

Suchmaschinen und ähnliche Werkzeuge haben viel mit dem mathematischen Gebiet der Booleschen Logik zu tun. Dieser Bereich befasst sich mit Aussagen, die nur wahr oder falsch sein können, andere Möglichkeiten gibt es nicht. So funktioniert ein Computer im Wesentlichen auch: 0 - 1, Strom - kein Strom. Alles andere entsteht durch bestimmte Kombinationen dieser Nullen und Einsen. Das ist sehr praktisch für die Suche: Entweder ein Wort ist im Text, oder eben nicht. Daraus kann man dann mit Verkettungen komplexere Aussagen basteln. Ohne ein Verständnis der folgenden Begriffe ist es bedeutend schwerer, sinnvolle Daten zu finden.

- *'logisch und'* bzw. *'AND'*

Bei einer Verkettung von zwei Stichworten mit AND werden nur Dateien zurückgeliefert, die beide Stichworte enthalten.

BSP: *'erich AND kästner'* findet nur Dokumente in denen beide Begriffe enthalten sind. Seiten nur mit *'erich'* werden ignoriert.

- *'logisch oder'* bzw. *'OR'*

Werden zwei Stichworte mit OR verbunden, dann beinhalten die zurückgelieferten Dateien mindestens eines der beiden Worte.

BSP: *'hund OR katze'* findet Dokumente, die das Wort Hund, das Wort Katze, oder beide Worte beinhalten.

- *'Negation'* bzw. *'NOT'*

Bedeutet, dass das nachfolgende Stichwort nicht im Text sein darf.

BSP: *'hund NOT katze'* findet Ergebnisse, die das Wort Hund enthalten, das Wort Katze darf aber nicht stehen.

- Verkettung

Aus diesen Elementen können größere Aussagen zusammengesetzt werden. Dazu sind Klammern nötig.

BSP: *'katze OR (hund NOT dackel)'* bedeutet, dass Dokumente mit dem Wort *'katze'* angezeigt werden. Dateien, die das Wort *'hund'* enthalten, werden nur angezeigt, wenn *'dackel'* nicht in ihnen vorkommt.

Das vorangestellte *'logisch'* soll hervorheben, dass es sich hierbei nicht um die umgangssprachlichen Begriffe handelt. Wenn wir im Deutschen nach Kaffee oder Tee fragen, dann ist die Antwort *'Beides, bitte!'* nicht die Erwartete, es soll sich nämlich zwischen den beiden Möglichkeiten entschieden werden. Bei *logisch oder* dagegen können wir problemlos beides kriegen.

2.1 Allgemeine Tipps

- Der wichtigste Punkt: Nicht direkt nach dem suchen, was man will, sondern nach dem, was höchstwahrscheinlich auf der Seite stehen könnte.

BSP: Bei der Suche nach einer deutschen Seite nicht 'deutsch' angeben, denn es ist sehr unwahrscheinlich, dass 'Diese Seite wurde in deutscher Sprache verfasst' im Dokument steht. Besser: '+und'. Warum? 'und' ist das zweithäufigste deutsche Wort (das häufigste ist 'die', aber das gibt es auch in anderen Sprachen). Die Wahrscheinlichkeit, dass es in unserem Text auftaucht, ist also sehr hoch.

- Vor dem Aufbau der Anfrage sollte man sich klar werden, auf welcher Art von Seiten die gesuchten Daten aufzufinden sind. Wenn das Ziel Online-Versionen von Zeitungen sind, dann wird dort eher unspezifische Sprache verwendet, während bei einer Suche nach spezialisierten Dokumenten umschreibende Fachworte sinnvoller sind.²
- Bei technischen oder im weitesten Sinne naturwissenschaftlichen Begriffen hilft auch oft die Verwendung des Wortes 'FAQ', es steht für Frequently Asked Questions. Das ist eine Liste mit zum Thema häufig gestellten Fragen.
- Begriffe wie FAQ heißen Akronyme (engl. acronym), also Abkürzungen. Bei einer ungeklärten Abkürzung lässt sich gut nach 'Suchbegriff acronym' suchen, also zum Beispiel 'faq acronym'. Dadurch findet man in der Regel Listen mit Erklärungen der Akronyme. Noch bessere Suchergebnisse können erzielt werden, wenn bereits andere Abkürzungen bekannt sind, am besten mit dem gleichen Anfangsbuchstaben (Dann stehen sie auch in einer nach Buchstaben eingeteilten Liste im gleichen Dokument).
- Sinnvoll [sollte getestet werden, bei Google und AlltheWeb klappt es nicht] kann es auch sein, die Wortendung nicht anzugeben,

BSP: 'crash' statt 'crashes'.

Dadurch beinhaltet die Ergebnismenge alle Dokumente, die 'crash', 'crashes', 'crashed', usw. enthalten, weil 'crash' ja in jedem diese Worte vorkommt. Manchmal kann dafür der Asterix '*' verwendet werden:

BSP: 'crash*'

- Generell ist es empfehlenswert, mit mehreren Suchmaschinen und Katalogen zu arbeiten, als nicht nur Google, sondern z.B. auch mit Altavista und AlltheWeb. Keine der Suchmaschinen hat ein auch nur ansatzweise komplettes Archiv. Außerdem haben verschiedene Suchmaschinen eigene geopolitische Schwerpunkte: Google z.B. hat seinen Themenbezug in Nordamerika (z.B. gut an der Nachrichtensuche erkennbar), AlltheWeb in Europa. Falls auf einer Seite eine lokale Suchmaschine benutzt werden kann, dann dürfte sie in der Regel bessere Ergebnisse als die globalen liefern.
- Wer genauere Informationen über die jeweilige Suchmaschine wünscht, kann sich praktisch immer auf deren Homepages tiefergehend informieren. Teilweise gibt es direkt Links im Anhang.

3 Kataloge

Als Kataloge oder Verzeichnisse werden Projekte bezeichnet, in denen Menschen Webseiten thematisch kategorisieren. Ein Beispiel für einen Katalog ist Yahoo. Früher war es bei Yahoo so, dass für diese Firma

²Die Suche nach Artikeln in Zeitschriften kann also auch genutzt werden, um Fachworte zu erfahren, mit denen dann die wirklich praktischen Artikel gefunden werden!

angestellte Menschen durch das Netz browsen und die gefundenen Seiten in das Verzeichnis eintragen. Mit zunehmender Bekanntheit des Dienstes kamen immer mehr Menschen auf die Idee, ihre Seiten aktiv bei Yahoo anzumelden, damit sie eingetragen werden. Bald waren die bei Yahoo Arbeitenden praktisch nur damit beschäftigt, diese Anfragen zu kategorisieren.

Bei Katalogen verspricht man sich bessere Suchergebnisse, da die Eintragung nicht automatisiert, sondern von Menschenhand stattfindet. Der Nachteil ist die geringere Geschwindigkeit und damit die potentiell geringere Aktualität. Außerdem unterliegen die Einträge damit menschlicher Willkür (Das ist bei Suchmaschinen ähnlich, aber nicht so offensichtlich).

3.1 Yahoo

In der Standardanfrage werden Worte mit OR verknüpft. Ist es erwünscht, dass die Worte alle vorkommen, dann sollte ein '+' direkt vor das jeweilige Wort geschrieben werden.

BSP: '+medien +bildung'

Bei der Suche nach einer Wortgruppe können Anführungsstriche verwendet werden.

BSP: '"suchen einer wortgruppe"'

Ausschließen eines Begriffes (NOT) funktioniert auch genau wie bei den meisten Suchwerkzeugen:

BSP: '+medien -bildung'

Mehr Funktionen lassen sich nicht ohne weiteres erfahren, in der Yahoo-Hilfe wird nur darauf verwiesen, dass das manuelle Durchstöbern der Verzeichnisse empfohlen wird.

3.2 The Virtual Library

Die WWW Virtual Library ist ein freies Projekt, gegründet 1991 von Tim Berners-Lee, dem Erfinder des World Wide Web. Hier gibt es verschiedene Freiwillige, die in bestimmten Bereichen kompetent sind, und eben diese Bereiche im Katalog warten.

Bei der Suchanfrage kann zwischen 'Any', 'All' und 'Boolean' gewählt werden. 'Any' bedeutet logisch oder, es muss also mindestens einer der gesuchten Begriffe im Text vorhanden sein. Bei 'All' handelt es sich um die Suche mit AND, es müssen also in der Ergebnisseite alle Suchbegriffe auftauchen. Wird die letzte Auswahlmöglichkeit genommen, dann können komplexere logische Verknüpfungen mit AND, OR, NOT und Klammern zusammengesetzt werden.

BSP: 'katze OR (hund NOT dackel)' bedeutet, dass das Ergebnis entweder das Wort 'katze' oder 'hund', bei 'hund' aber nicht 'dackel' enthalten soll.

Wie bei Katalogen üblich wird auch hier empfohlen, sich manuell durch die Verzeichnisstruktur zu hangeln, um an relevante Informationen zu gelangen.

3.3 dmoz open directory project

Dieser Katalog hat von seinen Ausmaßen her alle anderen Projekte dieser Art hinter sich gelassen. Auch hier handelt es sich um ein freies Projekt, das also von der freiwilligen und unbezahlten Teilnahme unzähliger Menschen lebt. Dabei sind sie so erfolgreich, dass viele große Anbieter wie AOL, Google und Lycos dmoz nutzen, um selber Katalog-Dienste anzubieten. Wenn man also z.B. den Katalog von Lycos nutzt, dann befindet man sich tatsächlich in den Archiven vom Open Directory Project.

Dmoz ist eines der wenigen großen Verzeichnisse, das komplexere Suchanfragen ermöglicht. Wie normalerweise üblich werden auch hier die gesuchten Begriffe mit AND verknüpft. Unerwünschte Worte können durch 'andnot' bzw. '-' ausgeschlossen werden. Nach 'and' und 'or' kann nicht gesucht werden, da diese als Boolesche Operatoren reserviert sind. Schön ist, dass hier auch die Suche mit einem Asterix klappt:

BSP: 'fahr*' liefert 'fahren', 'fahrer', 'fahrrad', ...

Das klappt aber nur am Ende des Wortes, woanders kann der Asterix nicht gesetzt werden. Außerdem kann festgelegt werden, ob der Suchbegriff im Titel ('t:'), in der Beschreibung ('d:' für description) oder in der URL ('u:') vorhanden ist.

BSP: 'd:overview t:basketball u:teams' sucht nach einem Dokument, in dessen Titel etwas mit Basketball steht, wobei in der URL das Wort teams vorhanden ist. Außerdem soll die Beschreibung der Seite das Wort overview enthalten.

4 Suchmaschinen

Im Gegensatz zu Verzeichnissen/Katalogen bauen Suchmaschinen ihre Datenbank dadurch auf, dass sie automatische Browser (Bots/Spider, das sind die eigentlichen Suchmaschinen, die Webseite stellt nur einen Zugriff auf die Datenbank bereit) durch das Netz schicken und die gefundenen Daten ebenfalls automatisch in die Datenbank eintragen. Diese Methode ist bedeutend schneller als die der Kataloge, dafür aber auch fehleranfälliger, was die Treffergenauigkeit angeht (So ist es manchmal empfehlenswert, bei der Anfrage mit Hilfe der Negation Begriffe wie 'xxx' oder 'sex' zu filtern, weil viele Erotikanbieter die abwegigsten technischen Stichworte in ihre Metatags einfügen).

4.1 Google

Google ist zur Zeit(!) die Suchmaschine mit der umfassendsten Datensammlung. Auch ist die Datenbank relativ treffsicher, weil die Anzahl der Begriffe, die in Metatags stehen dürfen, auf ein geringes Maß beschränkt wurde. Alles, was darüber liegt, wird als unglaubwürdig eingestuft und ignoriert.

In der Standardanfrage bei Google werden die eingegebenen Begriffe mit 'AND' verknüpft, es kommen also alle Worte in den Ergebnissen vor.

BSP: 'handball tennis' bringt nur Seiten, die sowohl 'handball', als auch 'tennis' enthalten.

Ob Worte groß oder klein geschrieben wurden, ist bei einer Google-Anfrage nicht von Relevanz.

BSP: 'handball tennis' und 'HandBall TENNIS' liefern die gleichen Ergebnisse.

Außerdem spart Google Ressourcen, indem besonders häufige Worte ignoriert werden. Wenn also Worte wie of, to, und, it, ... speziell erwünscht sind, dann ist es möglich, ein '+' als Präfix einzusetzen.

BSP: `"drum +and bass"`

Hier sieht man bereits ein weiteres Merkmal, nämlich die Möglichkeit, Wortgruppen zusammenhängend verarbeiten zu lassen. Dafür kann man sie in Anführungsstriche setzen. Wenn wir Daten zur Musikrichtung Drum & Bass suchen, aber keine Anführungsstriche setzen, dann werden alle Dokumente zurückgeliefert, die beide Worte *irgendwo im Text* enthalten. Das Ergebnis wäre vermutlich eine Liste mit Anbietern von Schlagzeugen und Zubehör. Die Anführungsstriche sind bei Wortgruppen also wesentlich.

Auch ist es mit Google möglich, die Suche auf bestimmte Domains zu beschränken. Möglich sind hier sowohl Top-Level-Domains wie 'de', als auch Domains wie 'uni-rostock.de' und sogar Sub-Domains, z.B. 'www.wiwi.uni-rostock.de'. Diese Eingrenzung erfolgt mit Hilfe des Begriffes 'site:'.

BSP: `'site:uni-rostock.de fernstudium'`

Dies ist sehr sinnvoll, wenn man sich nicht durch dutzende Menüs hangeln möchte, oder z.B. die Seite unübersichtlich designt wurde.

Unter Umständen ist es erwünscht, Begriffe aus der Suche auszuschließen, weil von vornherein bekannt ist, dass diese Begriffe nichts mit den gesuchten Daten zu tun haben oder einfach unerwünscht sind.

BSP: `"web browser" windows -IE -"Internet Explorer"`

Mit dieser Funktion muss aber aufgepasst werden, da damit z.B. Auflistungen zur Thematik nicht angezeigt werden, wenn der negierte Begriff enthalten ist.

Die Möglichkeit für komplexere Anfragen gibt es bei der "erweiterten Suche" oder "advanced search". Hier ist es z.B. möglich, nach bestimmten Dateierendungen oder nur nach Informationen neueren Datums zu suchen. Auch komplexere Verkettungen mit OR, NOT, usw. lassen sich hier komfortabler gestalten.

4.1.1 Googles Filter

Was bei Google noch einer Erwähnung bedarf sind die Zensurmechanismen. Der erste der beiden hat das Ziel, jugendgefährdende Daten zu filtern. Da die Filterung automatisch geschieht, kann davon ausgegangen werden, dass sie fehlerhaft ist. Um sie auszuschalten kann bei bereits eingegebener Suchanfrage an die von Google zurückgelieferte URL noch ein "&safe=off" angehängt und mit der Enter-Taste bestätigt werden.

BSP: aus

`"http://www.google.com/search?q=foo&start=0&start=0&ie=utf-8&oe=utf-8"`

wird

`"http://www.google.com/search?q=foo&start=0&start=0&ie=utf-8&oe=utf-8&safe=off"`

Es gibt auch noch die Möglichkeit, dies in den erweiterten Einstellungen von Google (oben rechts neben der Suchmaske) auszuwählen, aber diese Methode sollte, einmal gelernt, die Schnellere sein.

Google und andere Suchmaschinenbetreiber haben sich zu einer freiwilligen Selbstzensur zusammengeschlossen³, um sich rechtlich abzusichern. Das bedeutet für die Nutzenden, dass die Suchergebnisse in Deutschland anders aussehen als in Frankreich oder den USA. In letzterem Land ist diese Filterung verhältnismäßig gering⁴. Wenn aber versucht wird, auf Grund dieser Zensur <http://www.google.com> aufzurufen, so wird diese Anfrage nach z.B. <http://www.google.de> weitergeleitet. Wie kann man die erste gewählte Seite von der Weiterleitung abhalten? Die Lösung ist die folgende URL:

³http://www.tagesschau.de/aktuell/meldungen/0,1185,OID4100028_REF1,00.html

Google hat schon vorher zensiert, aber nun ist es amtlich.

⁴Es gibt im Netz Studien zum jeweiligen Ausmaß. Sie sollten nicht schwer zu finden sein ;>

Bsp: <http://www.google.com/ncr>

Das ncr steht für *no country relocation*, eine Weiterleitung auf die landesspezifische Seite (und ihre speziellen Filter) wird damit unterbunden. Letztlich ist damit natürlich das Problem der Zensur nicht aus der Welt, aber mensch ist wenigstens weiter arbeitsfähig.

4.2 AlltheWeb

Diese Suchmaschine hat ähnlich komplexe Fähigkeiten wie Google, auch die Syntax ist ähnlich. Hier haben z.B. die Anführungsstriche und + und - die gleichen Funktionen. Teilweise gibt es aber auch Unterschiede. So kann man bei AlltheWeb Begriffe mit OR verknüpfen, indem sie in Klammern gesetzt werden.

BSP: '(handball tennis)' bringt Dokumente, die mindestens eins der Worte enthalten.

Wenn nur Ergebnisse innerhalb einer bestimmten Domain gefragt sind, dann kann die Funktion site: benutzt werden.

BSP: 'site:uni-rostock.de fernstudium'

Diese kann natürlich auch ausschließend genutzt werden, falls zum Beispiel kommerzielle Angebote unerwünscht sind:

BSP: 'suchbegriff -site:com site:org' (Die Top Level Domain .com wurde für kommerzielle Anbieter im Netz erfunden.)

AlltheWeb bietet noch viele andere Funktionen, so z.B. eine Eingrenzung nach Dateigröße oder Dateityp. Bei Bedarf sei die vorbildliche Hilfe empfohlen.

4.3 Paperball

Für diejenigen, die auf der Suche nach Artikeln aus der Tagespresse sind, dürfte Paperball von Interesse sein. Paperball umfasst bei einer Suchanfrage zur Zeit über 200 Zeitungen. Die Suchanfrage muss bei Paperball etwas anders als bei Google gestellt werden. So macht es zum Beispiel einen Unterschied, ob Worte groß oder klein geschrieben werden.

BSP: 'liebe' liefert Dokumente, die sowohl 'Liebe', als auch das Verb 'liebe' enthalten.

Bei der Anfrage 'Liebe' dagegen werden nur Dokumente mit der großgeschriebenen Variante geliefert. Anfragen mit mehreren Begriffen werden als Standard mit OR verknüpft. Dies kann man mit '+' (AND) und '-'(NOT) direkt vor dem Suchwort umgehen.

BSP: 'handball tennis' bringt Seiten, die mindestens eins der Worte enthalten, '+handball -tennis' bringt nur Seiten, die auf jeden Fall 'handball', aber nicht 'tennis' enthalten.

Wortgruppen können wie üblich durch Anführungsstriche markiert werden.

Zusätzlich ist es möglich, in einzelnen Textfelder zu suchen, nämlich Titel (title), Vorspann (subtext), Text und Verfasser (author). Die englischen Ausdrücke sind hier auch angegeben, weil sie genau wie die deutschen genutzt werden können.

BSP: 'titel:irak'

5 Usenet

10 Jahre vor der Erfindung des World Wide Web entstand im Internet bereits ein Netz, das Usenet genannt wird. Hierbei handelt es sich sozusagen um ein gigantisches Forum. Tatsächlich sind die Foren, die im WWW oft anzutreffen sind, letztlich Kopien des Systems Usenet.⁵

Im Usenet gib es mehrere tausend solcher Foren, genannt Newsgroups, zu den verschiedensten Themen. Wenn man freundliche und kompetente Hilfe in diesen Foren erwartet, dann ist es sehr sinnvoll, sich an die Netiquette und ihre jeweiligen Ausprägungen der Newsgroup zu halten. Solche Informationen erhält man recht schnell, indem man die FAQ der jeweiligen Group liest bzw. einfach mal etwas mitliest, ohne sich einzumischen.

Das lohnt sich durchaus, denn im Usenet lassen sich viele kompetente Menschen mit den verschiedensten Interessengebieten finden. Mit einem Newsreader genannten Programm lässt sich auf das Usenet zugreifen (analog zum Browser für das WWW). Der Provider muss dies natürlich auch unterstützen, bei CallbyCall Anbietern ist dies äußerst unwahrscheinlich. Hier sei ein anderer Anbieter bzw. der kostenlose Dienst vom DFN unter <http://news.individual.de> empfohlen.

Beim Formulieren einer Anfrage ist es auch empfehlenswert, sich relativ kurz, aber präzise auszudrücken und evtl. auf andere fehlgeschlagene Versuche der Informationsbeschaffung zu verweisen. Je genauer das Problem geschildert wird, desto höher ist die Wahrscheinlichkeit einer sinnvollen Antwort. Wichtig ist: Eine Antwortgarantie gibt es nicht, es handelt sich hier um Menschen, die in ihrer Freizeit in den Groups kommunizieren, nicht um einen kostenlosen Auskunftsservice.

5.1 Google Groups

Google Groups ist ein Usenet-Archiv. Hier können praktisch alle Diskussionen, die in den letzten Jahren im Usenet stattfanden durchsucht werden. Die Wahrscheinlichkeit, dass die gesuchten Informationen bereits einmal im Usenet erfragt wurden, ist recht hoch, darum ist eine Suche in den Archiven sehr empfehlenswert. Die Syntax ist die gleiche wie bei der klassischen Google-Suchmaschine.

6 Hilfsmittel

6.1 Mailinglisten

Durch die zunehmende Überflutung des Usenets mit Spam und Newbies (im negativen Sinne) haben sich Mailinglisten als eine bedeutende Alternative entwickelt. Es gibt sie mittlerweile in ähnlich großer Vielfalt wie die Newsgroups im Usenet.

Mailinglisten funktionieren wie folgt: Es gibt auf einem Server eine Liste mit Emailadressen. Wenn jemand an diesen Server nun eine Mail schickt, dann verteilt der Server diese an alle, die in der Liste stehen. Mit Antworten auf die Mail wird ebenso verfahren. Dadurch wurde eine geschlossene Gruppe definiert, in der Kommunikation stattfinden kann. Und diese Möglichkeit wird für eine sehr breitgefächerte Menge von Thematiken angeboten. Mailinglisten, die nur in eine Richtung funktionieren (vom Server zu den Adressaten), werden auch Newsletter genannt.

Oft gibt es von den in Mailinglisten stattfindenden Diskussionen Archive im Web. Für weniger komplexe Fragen ist dies dann meist die beste Option.

Wenn umfassendere Probleme behandelt werden sollen bzw. wenn weitergehende Kommunikation zur Materie angestrebt wird, dann kann man sich in die Mailinglisten eintragen (Sie haben meist auch eine Website mit Erklärungen). Das Diskussionsklima ist in der Regel ein sehr freundliches, da jemand, der sich nicht an die Netiquette hält, einfach vom Administrator von der Liste verbannt wird.

⁵Natürlich ist niemand gezwungen, die Inhalte dieses Dokumentes einfach so zu glauben, ruhig mal zum Testen der eigenen Fähigkeiten eine Suchanfrage starten!

6.2 Web Archive

Ein Charakteristikum des WWW ist die ständige Fluktuation der Seiten. Falls ein Ergebnis gefunden wurde, das nur eine 404 Fehlermeldung hervorbrachte, von dem aber vermutet wird, dass es sich hierbei genau um das gesuchte Dokument handelt, dann kann das Web Archive die entscheidende Hilfe darstellen. Dieses Projekt hat die Zielstellung, ältere oder nicht mehr vorhandene Seiten zu archivieren. Bei Eingabe einer Domain/Webseite kann die jeweilige Version in Monatsabständen betrachtet werden.

Auch die Suchmaschine Google besitzt einen Cache der gescannten Seiten. Benutzbar ist er, wenn neben dem gefundenen Dokument ein mit 'cached' oder 'Im Archiv' bezeichneter Link vorhanden ist.

A Anhang

A.1 Suchen für Fortgeschrittene

Das Gebiet der Informationssuche ist ein äußerst komplexes, darum sei an dieser Stelle nur auf das exzellente Projekt Searchlores.org verwiesen. Diese Seiten umfasst eine sehr umfangreiche Sammlung von Essays zu verschiedensten Aspekten der Thematik, man könnte mit der Datenmenge eine ganze Bibliothek füllen.⁶ Dieser Text und Searchlores sollten mehr als ausreichen, um Interessierte zu befähigen, von nun an allein zu suchen.

A.2 Software

Im Folgenden werden einige freie Softwarepakete vorgestellt, mit denen selbst unter Windows die Netznutzung recht angenehm wird. Von den jeweiligen Microsoftprodukten wird aus Sicherheits- sowie Effizienzgründen abgeraten.

A.2.1 Browser: Firefox

Firefox ist ein Abkömmling der Mozilla-Entwicklung. Im Gegensatz zu Mozilla handelt es sich bei Firefox nur um einen Browser, kein Paket mit Mailtool und Newsreader. Projektziele von Firefox sind geringere Größe, höhere Geschwindigkeit und einfacher strukturierte Oberfläche. Der Browser hat viele Optionen, die die Netznutzung erleichtern. Es gibt z.B. ein Eingabefeld neben der Adresseingabe, das direkt die eingegebenen Begriffe bei Google suchen lässt. Bei häufiger Nutzung von Google macht sich der geringere Suchaufwand deutlich bemerkbar. Firefox lässt sich recht stark an die persönlichen Bedürfnisse anpassen, sollte also auch für höhere Ansprüche genügen.

Die einzige qualitativ gleichwertige Alternative ist Opera (Das gilt nur für die Versionen vor 6, alles darüber muss als Bloatware bezeichnet werden), dieser Browser hat die gleichen Fähigkeiten und ist zudem kleiner und schneller, allerdings in der kostenlosen Version mit einem Werbebanner versehen. Zudem ist dieses Programm nicht frei.

A.2.2 Mailtool + Newsreader: Sylpheed

Dieses Programm hat alle Features, die man von einem fortschrittlichen Mailtool erwarten darf: eine intuitive Oberfläche, volle Unterstützung von Verschlüsselung der Verbindung und Unterstützung von Verschlüsselung für die Mails selbst, Nutzung mehrerer Accounts, ...

Fortgeschrittenere User werden sich bei Sylpheed über die Möglichkeit freuen, beinahe alles mit der Tastatur steuern zu können.

⁶Allerdings wäre das sinnfrei, gibt es die Informationen doch im Web ;-)

Dieses Tool wurde ursprünglich für Unix-ähnliche Betriebssysteme entwickelt, aber unter der in den Links stehenden genannten Adresse ist eine Version erhältlich, die auch auf Windows läuft.

Die Version 'sylpheed-claws complete installer' wird empfohlen, da sie sich leicht installieren lässt und am aktuellsten ist.

A.3 Glossar

404:	Eine Fehlermeldung von Webservern, wenn ein Dokument nicht gefunden wurde.
Bloatware:	Software, aus der die Entwickler im ständigen Druck, neue Fähigkeiten implementieren zu müssen, ein viel Speicher und Prozessorleistung fressendes System geschaffen haben, das auch mit der Hälfte des Ressourcenverbrauchs praktisch alle Fähigkeiten haben könnte. Ein tolles Beispiel ist z.B der Microsoft Internet Explorer: 80 MB gegen 4MB für Opera 5.
Bots:	siehe Robots
Forum:	Auf einer Webseite können Nachrichten geschrieben werden, die dann für die Öffentlichkeit les- und kommentierbar sind.
Metatags:	Autoren von Webseiten können Stichworte, die den Inhalt der Seite betreffen sollen, in den Quelltext einfügen. Diese sind auf der Seite selbst nicht sichtbar, werden aber (meist) von Suchmaschinen verarbeitet.
Mozilla:	Netscape hat den Quellcode für den eigenen Browser Communicator in die Öffentlichkeit gestellt, also ein Open Source Projekt begründet. Dieses nennt sich Mozilla.
Netiquette:	Eine Sammlung von (ursprünglich) ungeschriebenen Verhaltensregeln für das Internet.
Newbie:	Ein unerfahrener User. Kann je nach Auftreten desselben freundlich, aber auch negativ gemeint sein.
Open Source:	Programme werden meist nicht in Maschinensprache, also Einsen und Nullen, sondern in leichter verständlichen Programmiersprachen geschrieben. Ein spezielles Programm übersetzt das dann in Maschinensprache, die der Computer versteht. Die in dieser Programmiersprache geschriebenen Texte heißen Quelltext (engl. source code). Bei kommerziellen Programmen sind diese Quelltexte nicht erhältlich, und die Weitergabe ist verboten (damit soll schließlich Geld gemacht werden). Open source (genauer: freie Software ⁷) bedeutet, dass der Quelltext für alle ersichtlich ist, und alle Menschen das Recht haben, ihn zu benutzen, zu vervielfältigen und zu verändern.
Robots:	Automatisch arbeitende Programme, die Webseiten besuchen und die Inhalte in Datenbanken schreiben.
Server:	Ein Computer, der Dienste wie z.B. Mail anbietet (server ist das englische Wort für Diener).
Spam:	Massenmails, meist mit der Zielsetzung, die empfangende Person vom Kauf eines in der Mail beworbenen Objektes zu überzeugen.

⁷Es gibt Unterschiede zwischen Open Source und freier Software, die sich an den Rechten festmachen, die den UserInnen ermöglicht werden. Aus deren Sicht ist freie Software zu bevorzugen, allerdings hat der Begriff Open Source einfach mehr Appeal, was in der Öffentlichkeit zu einer Verwischung führt. Bei Interesse sollte eine Anfrage mit den Stichworten 'open source', 'free software' und 'differences' den Wissensdurst stillen.

Spiders:	siehe Robots
Top Level Domain	(TLD): die Endungen einer Adresse, also die zweibuchstabigen Länderendungen wie de oder die Domains com, net, org usw.
UNIX:	Ein Konzept eines Betriebssystems, die Anfänge liegen in den 70ern. Praktisch alle aktuellen, professionellen Betriebssysteme sind recht UNIX-ähnlich.
URL:	Unified Resource Locator. Auch als Internet-Adresse bekannt. Ein Beispiel: http://www.debian.de/
Web:	siehe World Wide Web
Webserver:	Ein Server, der Html-Dokumente über das Web schickt, wenn man sie mit einem Browser anfragt.
WWW:	siehe Word Wide Web
World Wide Web:	Einer der Dienste im Internet. Wohlgermerkt ist das Web nicht das gleiche wie das Internet. Das Protokoll des WWW heißt Hypertext Transfer Protokoll(HTTP).

A.4 Links

AlltheWeb:	http://www.alltheweb.com http://www.alltheweb.com
AlltheWeb Hilfe:	http://www.alltheweb.com/help/faqs/web_search.html
DFN NetNews:	http://news.individual.de/
dmoz:	http://www.dmoz.org/
Firefox:	http://www.mozilla.org/products/firefox/
Google:	http://www.google.com/
Google Hilfe:	http://www.google.de/intl/de/help.html
Google Groups:	http://groups.google.com/
Opera:	http://www.opera.com/
Paperball:	http://www.paperball.de
Paperball Hilfe:	http://paperball.fireball.de/paperball_hilfe.html
Searchlores:	http://www.searchlore.org/ oder http://www.searchlores.org/
Sylpheed für Windows	http://sylpheed-claws.sourceforge.net/win32/ Download unter 'Sylpheed Claws für Windows'
The Virtual Library:	http://vlib.org/
WayBack Machine:	http://www.archive.org/
Web Archive:	http://webdev.archive.org/
Yahoo:	http://www.yahoo.com/

A.5 Lizenz

Dieses Werk untersteht der Lizenz *Creative Commons NonCommercial-ShareAlike 1.0*. Dies heißt, daß Du außer den üblichen Rechten zusätzlich noch folgendes darfst:

- dieses Werk kopieren, verteilen, darstellen und daraus andere Arbeiten ableiten,

solange Du dabei folgendes beachtest:

- Dieses Werk darf nicht für kommerzielle Zwecke genutzt werden.
- Wenn Du dieses Werk veränderst oder Deine Arbeit darauf aufbaust, dann muss das resultierende Werk auch sichtbar unter dieser Lizenz stehen.

Die gleiche Aussage noch einmal in einer JuristInnen-kompatiblen Sprache:
<http://creativecommons.org/licenses/nc-sa/1.0/legalcode>