

Recherchieren im Netz - Dos & Don'ts

Mit Tipps und Tricks durch den
Datendschungel von Google und Co.

Mag.(FH) Roman Huditsch
Arbeiterkammer Burgenland

Überblick

- Warum Suche im Internet?
- Tools für die Internetsuche
- Wie suche ich im Internet?
- Wie evaluiere ich Internetquellen?
- Quellenangaben

Warum Suche im Internet?

Vorteile der Internetsuche

- Informationen zu (fast) allen Informationsbedürfnissen
- Zugriff vom Arbeitsplatz aus
- Schnelle Ergebnisse
- Keine tiefgreifenden Vorkenntnisse notwendig
- Oft reicht eine Suchanfrage wirklich aus, um die gesuchte Information zu bekommen



Nachteile der Internetsuche

- Totales Überangebot an Information
- Wertungsfreies Nebeneinander von Qualität und Unsinn
- Optik in vielen Fällen besser als Inhalt
- Reihung der Suchergebnisse oft nicht nachvollziehbar
- Wird wirklich die beste Seite zum Thema gefunden?
- Sind die Informationen richtig und vollständig?
- Wie werte ich Informationen aus und wie ermittle ich, ob sie für meine Suche zutreffend sind?

Warum reicht es nicht nur Google zu benutzen?

- Über allgemeine Suchmaschinen (Google) auffindbar:
 - ca. 20 Milliarden Internetseiten
- Insgesamt im WWW (frei zugänglich):
 - mehr als 30 Milliarden Internetseiten
- Nicht über Suchmaschinen zu finden:
 - mind. 30 Milliarden Dokumente („Invisible Web“):
 - Nur ein Teil aller Informationen sind überhaupt in Suchmaschinen enthalten!



Was kann ich nicht finden?

- Frei zugängliche Internetseiten, die bei keinem Suchdienst gespeichert sind (z.B. Internetseiten, auf die keine Links führen)
 - Können von sog. Spidern nicht gefunden und indiziert werden
- Inhalte aus Datenbanken oder Bibliothekskatalogen („Invisible Web“)
- Geschützte Inhalte (nur mit Passwort zugänglich)
- Was es einfach nicht im Internet gibt (z.B. Copyright-geschütztes Material)

Grenzen der Recherche in Suchmaschinen

- „Grenzen“ des Suchenden
- Unvollständigkeit des Suchindexes
- Unbekannte Dateiformate
- Nicht-Text-Informationen
- Problem: Grafik- / Musikdateien ohne Beschreibungstexte
- Eingeschränkte Suchmöglichkeiten
- Inhaltliche Filterung / Zensur
- Veraltete Information („Seite nicht gefunden“)
- „Invisible Web“

Tools für die Internetsuche

Tools für die Internetsuche

- Internetverzeichnisse und Fachportale
- Suchmaschinen
- Spezialsuchmaschinen
- Metasuchmaschinen
- Suche im „Deep web“

Wann ist welches Tool sinnvoll?



Internetverzeichnisse

- = Intellektuell angelegte Linksammlungen
- für den Überblick auf ein Thema, zum Stöbern
- der Aktualitätsgrad ist zweitrangig
- Umfang: bis zu 4 Millionen Nachweise
- Akademische oder institutionelle Kataloge
 - geeignet für Forschungsinteressen
- Kommerzielle Kataloge
 - geeignet für persönliche Interessen

Beispiele für Internetverzeichnisse

- Kommerzielle Verzeichnisse:
 - Yahoo: de.yahoo.com/
 - Open Directory Project: <http://dmoz.org/>
- Akademische Verzeichnisse:
 - Internet Public Library: <http://www.ipl.org/>
 - The Virtual Library: www.vlib.org/
- Überblick:
 - www.bib-bvb.de/bib_schule/verzeich.pdf

Internet browser window showing the website **Deutsches Internet Verzeichnis** (DIV). The address bar shows <http://internet-verzeichnis.de/>.

The main navigation bar includes: **Deutsches Internet Verzeichnis**, **Firma kostenlos eintragen** (Firmenverzeichnis Eintrag in wenigen Minuten), and **Datenschutzinfo**.

Left sidebar (Hauptrubriken):

- Auskunft, Auto, Bildung, Computer, Essen, Freizeit, Gesundheit, Haus, Kunst, Nachrichten, Reisen, Sport, Wirtschaft
- Aktuelles: Wetter, Nachrichten

Search section (Suchen & Finden):

Suche in mehreren Suchmaschinen gleichzeitig >hier<

Enterprise Search Paket
HighEnd-Suche für Ihr Unternehmen. unternehmensinterne Suchmaschine!

Main content area (Home) with 14 Rubriken:

Rubrik	Anzahl Einträge	Beispiel
Auskunft	[11780]	Bankleitzahlen, Branchenführer, Branchenverzeichnis...
Bildung	[1643]	Bibliotheken, Bildungsmedien, Hilfen, Institute, Schulen...
Essen	[1971]	Diät, Ernährung, Fasten, Genußmittel, Getränke...
Gesellschaft	[2449]	Hilfsorganisationen, Organisationen...
Haus	[2913]	Garten, Haus...
Nachrichten	[389]	Fernsehen, Film, Freie Journalisten, Kino...
Sport	[665]	Angeln, Ausdauer, Ballsport, Basketball, Cheerleading...
Auto	[2208]	Anhänger, Anhängerkupplungen, Automobil-Industrie...
Computer	[1774]	Hardware, Software, Spiele...
Freizeit	[1315]	Bars, Cafes & Kneipen, Clubs & Discotheken, Film...
Gesundheit	[2126]	Alternativmedizin, Apotheken, Ernährung...
Kunst	[289]	Denkmale, Kulturgeschichte, Literatur Bücher, Museen...
Reisen	[4387]	Angelreisen, Autovermietungen, Bahnreisen, Ballonfahrt...
Wirtschaft	[12547]	Arbeit, Banken & Sparkassen, Branchen, Finanzamt...

Windows taskbar at the bottom shows the system clock: 15:22, 25.02.2015.

Vorteile von Internetverzeichnissen

- + **Ausgewählt:**
Internetseiten werden von Menschenhand ausgewählt und in ein passendes Themengebiet einsortiert
- + **Geprüft:**
Internetseiten gelangen nur nach vorheriger Prüfung in den Index
- + **Bewertet:**
Internetseiten sind teilweise „bewertet“

Nachteile von Internetverzeichnissen

- Geringerer Umfang:
Umfang < 0,05% im Vergleich zu Suchmaschinen
- Kleinerer Suchbereich:
i.d.R. sind nur der Titel und eine Kurzbeschreibung der Startseite eines Internetangebots suchbar
- fehlende Aktualität:
z.T. dauert es sehr lange, bis neue Internetseiten aufgenommen werden
- Fehlende Einheitlichkeit:
Struktur der Kategorien je nach Internetverzeichnis anders (kein System)

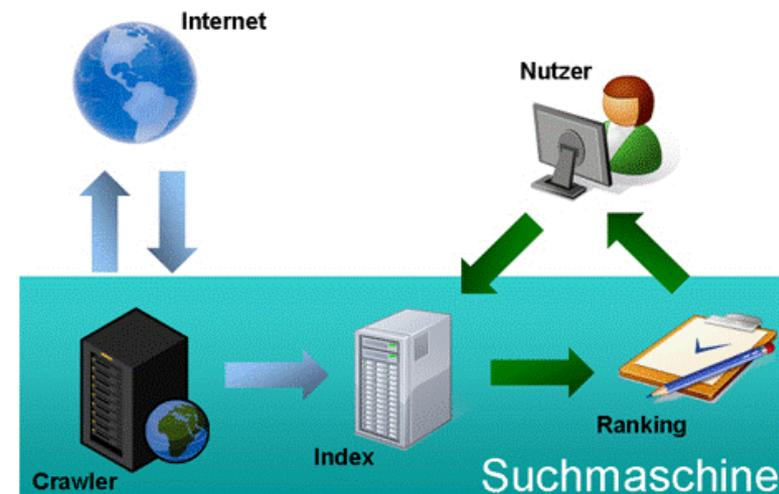
Suchmaschinen

- Suche nach Fakten oder aktuelleren Informationen
- indexieren möglichst alle Internetseiten eines Internetangebots vollständig, voll automatisch, (fast) Wort für Wort
- machen keinen Unterschied bei der Qualität der indexierten Internetseiten

The logo for Google Deutschland, featuring the word "Google" in its characteristic multi-colored font (blue, red, yellow, blue, green, red) and the word "Deutschland" in a smaller, grey font directly below it.

Wie funktionieren Suchmaschinen?

- Automatisch arbeitende Programme (Spider, Crawler): Hangeln sie sich über Verweise weiter
- Internetseiten werden indiziert, wobei der Volltext der Seiten in riesigen Datenbanken gespeichert wird
- Der Benutzer recherchiert in diesen Datenbanken durch Eingabe von Suchbegriffen
- Relevante Webseiten werden gereiht und als Ergebnismenge zurückgegeben



Suchtechnologie

- „First generation“
 - Automatisches Ranking: wie häufig und wo kommt das gesuchte Wort im Text vor?
- „Second generation“
 - Automatisches semantisches Ranking
 - „peer“ Ranking: die Zahl der Links auf das Dokument fließt mit ein
 - Concept clustering: Treffer werden thematisch geordnet
 - i.d.R. erhält man damit mehr relevante Treffer

Beispiele für Suchmaschinen

- Google: www.google.com
- Bing: www.bing.com
- Baidu: <http://www.baidu.com/>

Vorteile von Suchmaschinen

- + Schnell:
In Sekundenbruchteilen können mehrere Milliarden Seiten durchsucht werden
- + Relativ aktuell:
neue oder geänderte Internetseiten sind häufig schon nach wenigen Tagen im Index
- + Sehr umfangreich:
Internetseiten sind meist komplett gespeichert

Nachteile von Suchmaschinen

- Fehlende „Intelligenz“:
Intellektuelle Erschließung unmöglich, viel nutzlose „Information“ („Datenflut“)
- Nicht vollständig:
Suchmaschinen erreichen nur einen Teil aller Internetseiten weltweit
- Gemischte Aktualität:
Aktuelle Nachrichten stehen undifferenziert neben alten Daten
- „Suchmaschinenspam“:
Verfälschte Relevanzbewertung durch Link-Farmen

Spezielsuchmaschinen

- Informationen finden, die nicht / kaum durch allgemeine Suchmaschinen abgedeckt werden
- Umfang: Einige tausend bis zu 100 Millionen Internetseiten
- Weltweit gibt es Tausende von Suchdiensten, die auf einen bestimmten Bereich spezialisiert sind:
 - wissenschaftliche Informationen
 - tagesaktuelle Informationen (Tageszeitungen) – www.paperball.de
 - dauerhafte Archivierung von Webseiten
 - „Nicht-Text“, z.B. Grafiken, Filme oder Musik
 - Adressen, Telefonnummern, Firmen etc.

Beispiele für Spezialsuchmaschinen

- Scirus: <http://www.scirus.com>
 - Durchsucht Seiten von Universitäten, ausgewählte Server und wissenschaftliche Webseiten plus freie zugängliche Fachdatenbanken und Volltexte wissenschaftlicher Zeitschriften
- Human Protein Reference Database: <http://www.hprd.org/>
 - Eigenschaften von einzelnen Proteinen (Domänen, Modifikationen, Interaktionen, assoziierte Krankheitsbilder etc.) wurden manuell aus verschiedenen Literaturdatenbanken extrahiert

Vorteile von Spezialsuchmaschinen

- + Abdeckung:
Die „Randgebiete“ werden sehr viel besser abgedeckt
- + Eingrenzung:
Sie suchen nur in dem Bereich, der Sie interessiert
- + Relevanz:
oft hohe Relevanz der gefundenen Quellen

Nachteile von Spezialsuchmaschinen

- Such-Einschränkung:
meist nur für ein ganz bestimmtes Thema geeignet
- Schwer „aufzuspüren“:
passende „Spezialisten“ sind häufig schwierig aufzufinden
(Suchmaschinenverzeichnisse nutzen)
- Fehlende Einheitlichkeit:
Qualität und Umfang sehr unterschiedlich

Metasuchmaschinen

- Eingabe einer Suchanfrage bei einer Metasuchmaschine
- Metasuchmaschine gibt Anfrage an Suchdienste weiter
- Suchdienste suchen in ihrer Datenbank und liefern der Metasuchmaschine die Ergebnisse
- Ergebnisse werden von Metasuchmaschine gesammelt, zu einer Liste zusammengefügt und angezeigt - dabei werden doppelte Treffer teilweise entfernt oder zusammengefasst
- Beispiele
 - www.metacrawler.de
 - <http://www.iboogie.tv/>
 - www.ixquick.com

Deep Web

- = unsichtbares Netz (d.h. unsichtbar für Suchmaschinen):
- Beispiele
 - in Datenbanken gespeicherte Information (bahn.de, telefonbuch.de, Bibliothekskataloge, Beilstein, Scifinder...)
 - einige dynamische Seiten
 - andere Formate als Text

Weitere Informationsquellen

- Fachdatenbanken und Bibliothekskataloge:
 - Die Inhalte können i.d.R. nicht über allgemeine Suchmaschinen gefunden werden (Invisible Web)
- Nachschlagewerke (Wikipedia), Internet-Portale, Telefonbücher, Stadtpläne

Übung

- Wo gibt es die billigste Salami-Pizza in Nürnberg?
- Wie viele Bücher hat Martin Suter schon geschrieben und wie heißt sein neuestes?
- An welchem Film arbeitet Robert Downey JR. gerade?



Wie suche ich im Internet?

Suchkonzept und Suchstrategien

- Vor einer Recherche sollten Sie sich über einige Dinge klar werden, die dann zum Suchkonzept und zur geeigneten Suchstrategie führen
- Warum ?
 - Zur Strukturierung der eigenen Gedanken
 - Zur Begriffsklärung
 - Als Basis für die Auswahl der Rechercheinstrumente
 - Für das Zeitmanagement

Das Konzept

- Was suche ich überhaupt?
- Was sind die wichtigsten Konzepte meines Themas?
- Welche Suchbegriffe / Formeln liefert mir mein Konzept?
- Welche Synonyme, Abkürzungen und Flexionen sind möglich?
- Die gefundenen Begriffe können in einem nächsten Schritt miteinander verknüpft werden

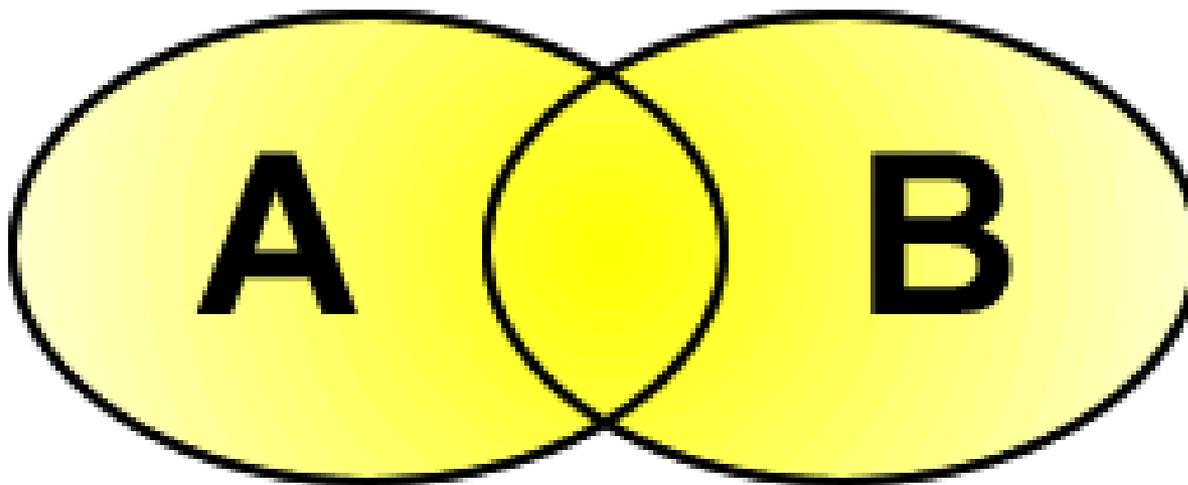
Bedarfsanalyse

- Wofür brauche ich die Information? Wie umfassend, wie genau, wie aktuell muss die Information sein?
 - Hausarbeit, Referat → ein paar gute Quellen
 - Abschlussarbeit im Anfangsstadium → umfassend (erster Überblick)
 - Abschlussarbeit in späteren Stadien, Forschung → genau, möglichst aktuell
- Danach richten sich Publikationstypen, Rechercheinstrumente, Suchbegriffe

Suchbegriffe kombinieren (ODER)

- Alle Treffer anzeigen, in denen mindestens eines der gesuchten Wörter vorkommt (Synonyme)

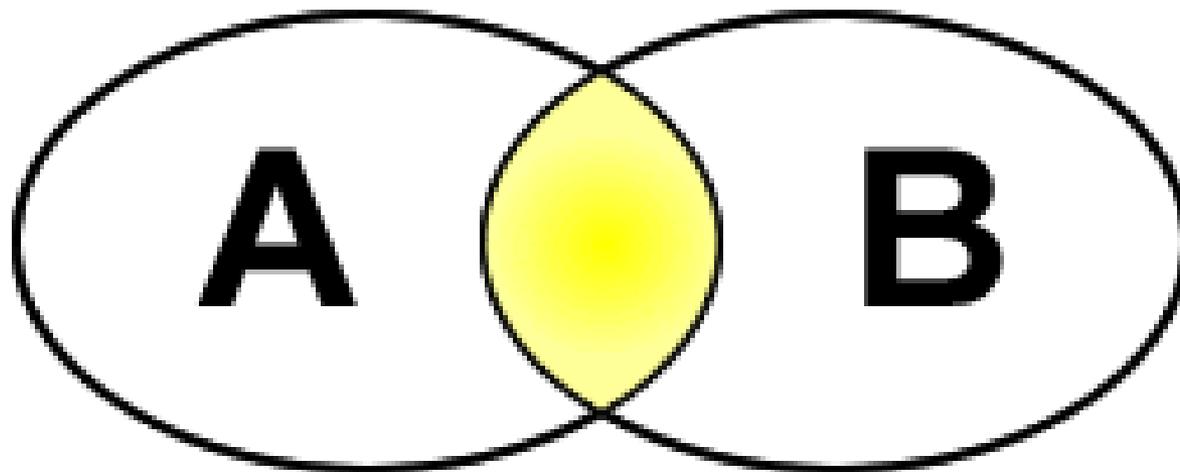
ferienwohnung OR ferienhaus



Suchbegriffe kombinieren (UND)

- Nur Treffer anzeigen, in denen alle gesuchten Wörter vorkommen (egal in welcher Reihenfolge)

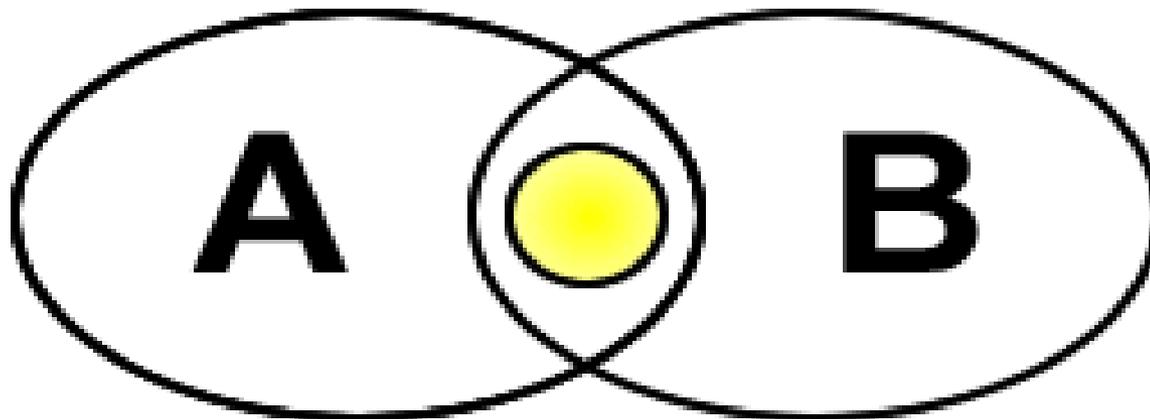
ferienwohnung AND ferienhaus



Suchbegriffe kombinieren (Phrase)

- Nur Treffer anzeigen, in denen alle gesuchten Wörter in exakt dieser Reihenfolge vorkommen

„Fachhochschule Eisenstadt“

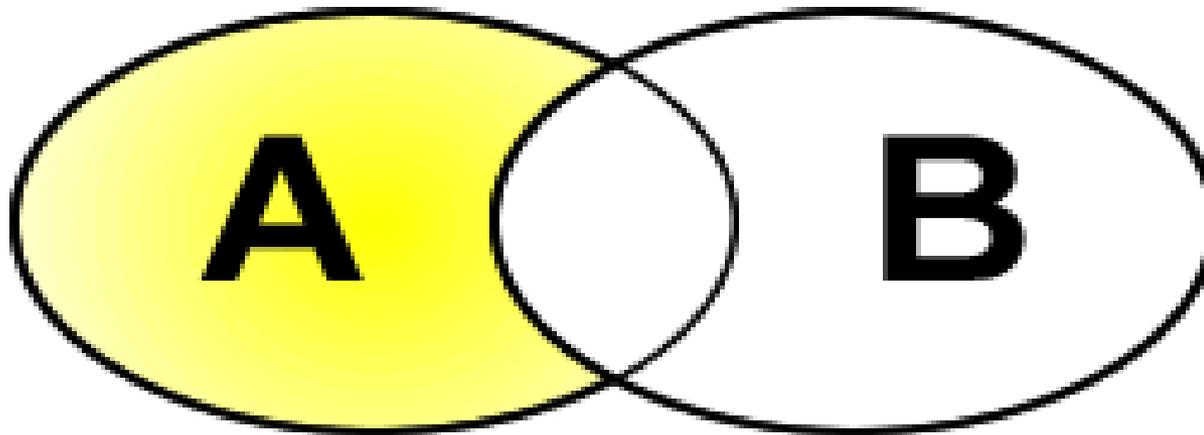


- Nicht gefunden wird z.B.: die Fachhochschule in Eisenstadt

Suchbegriffe kombinieren (NICHT)

- Nur die Treffer anzeigen, in denen ein bestimmtes Wort nicht vorkommt

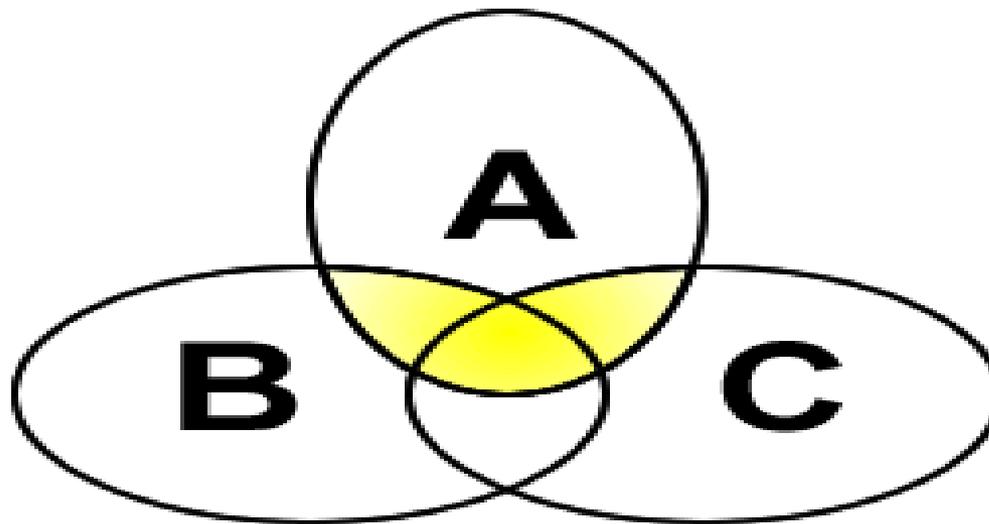
„Fachhochschule -Eisenstadt“



Suchbegriffe kombinieren (Synonyme)

- Nur Treffer anzeigen, in denen ein Wort vorkommen muss und mindestens ein weiteres der gesuchten Wörter (A+B oder A+C oder A+B+C)

„Fachhochschule (Eisenstadt OR Wiener Neustadt)“



Trunkieren

- Bei einer Trunkierung wird nach dem Wortstamm (dem sogenannten "trunk") gesucht. Man gibt einfach nur einen Teil des Wortes ein und setzt dann ein Trunkierungszeichen.
- * oder ?

„Fachhochschul* (Eisenstadt OR Wiener Neustadt)“

Recherche modifizieren - Zu wenig Treffer

- Relevante Treffer sichten
 - Suche nach weiteren Synonymen/Schlagwörtern
- Thema weiter fassen
 - Wortliste ergänzen
 - mittels Trunkierung erweitern
- Fehlerquellensuchen
 - Tippfehler
 - falsche Platzhalter
 - Falsche Suchlogik
 - Voreinstellungen nicht beachtet?

Recherche modifizieren - Zu viele Treffer

- Relevante Treffer sichten
 - Suche nach genaueren Synonymen/Schlagwörtern
- Thema enger fassen
 - Nebenaspekte weg lassen
 - Trunkierungen einschränken
- Irrelevante Treffer sichten
 - Arbeiten mit dem Nicht-Operator
- **Speziellere Datenbanken benutzen**
- In der Bibliothek fragen

Wie evaluiere ich Internetquellen?

Beurteilung einer Internetseite – Warum?

- Generell sollte jede Informationsquelle kritisch hinterfragt werden
- „Problem“: Keine Kontrollinstanz im Web
 - jeder kann Inhalte veröffentlichen
 - Suchmaschinen indizieren jeden Text
- Neben vielen seriösen Informationen:
 - Informationen, deren Wahrheitsgehalt schwer zu prüfen ist
 - Internetseiten, die bewusst oder unbewusst fehlerhafte oder verfälschte Inhalte enthalten

Formale Kriterien zur Beurteilung einer

- Wer betreibt das Internetangebot?
 - wissenschaftliche Einrichtung, Firma, Privatperson?
 - Impressum, Link auf die Einstiegsseite
- Wer ist der Verfasser und in welcher Funktion schreibt er?
 - Angabe am Fuß einer Seite
 - „Wir über uns“, Biographie
- Wie aktuell sind die Informationen?
 - Letztes Aktualisierungsdatum
 - Angaben im Inhalt
 - Anteil der funktionierenden Links

Inhaltliche Kriterien zur Beurteilung einer

- Für wen wurde die Seite geschrieben?
 - Allgemeinheit, Schüler, Studierende, Wissenschaftler?
 - Niveau, Umfang, Verwendung von Fachbegriffen und Fremdwörtern
 - Qualität der externen Links
- Wie ist der Autor zu seiner Aussage gekommen?
 - Sind Aussagen, Behauptungen, Thesen durch einschlägige Quellen belegt?
- Gibt es Kommentare von anderen Nutzern?
 - Hinweis auf Unstimmigkeiten, Fehler
 - Zustimmung oder Anerkennung

Kriterien zur Beurteilung einer

- Eine „passende“ Internetseite als Ausgangspunkt nutzen:
 - Links auf der Seite folgen
 - Einschlägige Begriffe oder Namen (Autoren) als neue Suchbegriffe in einer Suchmaschine nutzen
 - „Linksuche“ in Suchmaschinen, um thematisch ähnliche Seiten zu finden

Beurteilung der Qualität einer Trefferliste

- Eine „passende“ Internetseite als Ausgangspunkt nutzen:
 - Links auf der Seite folgen
 - Einschlägige Begriffe oder Namen (Autoren) als neue Suchbegriffe in einer Suchmaschine nutzen
 - „Linksuche“ in Suchmaschinen, um thematisch ähnliche Seiten zu finden

Quellenangaben

Quellenangaben

- www.ub.uni-konstanz.de/fileadmin/Dateien/...6/6_internet.ppt
- www.uni-saarland.de/uploads/media/Lit.-Recherche_f._Fortgeschr.ppt
- www.informationskompetenz.de/.../Suchen%20im%20Internet_583.ppt