

Informationssuche in der Bibliothek und im Internet (WWW)

- *Wie finde ich Information, die nicht im Schulbuch steht?* -

Kernaufgaben: *Informationen finden und bewerten*

1. Wie finde ich Information in der **Bibliothek**: Bücherarten und **Stellordnung**
2. Wie finde ich Information im WWW: **Suchtechnik** in Datenbanken und im Web

Übungsmodelle von Datenbanken: - Katalogdatenbank von BIBLIOTHECA2000
 - **Google - erweiterte Suche** (+ Wikipedia)
 - JUFOBASE (Jugend Forscht-Sieger-Arbeiten)

1. Suchweg: **Die Bibliotheken**

A. **Holzschuh-Bibliothek**

Nutzt das System der Stellordnung ASB !

Suche in der Katalogdatenbank oder im „**OPAC**“, dem Benutzerterminal von Bibliotheca2000 (inkl. Biologie- und Chemiebücher der Lehrer).

=> **die Katalogdatenbank ist geeignet, die Datenbankrecherche zu üben**

B. **Gemeindebibliothek Karlsbad und Stadtbibliothek Ettlingen:**

Verwenden gleiches System: Suche wird im „**Web-OPAC**“ ausgeführt.
Ein ähnliches System verwendet die **Stadtbibliothek Karlsruhe**.

Wichtig: Diese Suchwege finden **nur Bücher, CDs und Ähnliches, keine Fachaufsätze**.

2. Suchweg: **Das World Wide Web im Internet**

Web-Dokumente sind meist keine abgeschlossenen Werke, sondern haben den Charakter von Lexikon- oder Zeitschriftenartikeln. Nützlich sind **Einstiegsseiten**, die mit eigenem Suchsystem und eigenem Linkbestand den **Fachliteraturgehalt des WWW** erschließen. Vorsicht vor „Wegweisern zu verschlossenen Türen“.

Für viele Web-Angebote ist das **Netz nur Transportmedium**. Wertvolle Information ist meist kostenpflichtig.

Leitgedanken:

- Was machen Suchmaschinen eigentlich?
- Was gibt es noch, außer **Google** und **Wikipedia**?
- Die Vielfalt ist riesig, aber endlich.
- **Nicht immer mit dem ersten Treffer zufrieden sein.**

Schwerpunkte:

- **Allgemeine Suchmaschinen:** Beispiel Google
- **Dokumentsammlungen:** Beispiel Wikipedia und Alternativen
- Ausblick: „**Fachportale**“: Informationsquellen für bestimmte Fachgebiete, z. B. **vifabio** (Biologie)
umfassende **Bibliothekskataloge:** **Karlsruher Virtueller Katalog (KVK), VASCODA**

A. Die Informationswelt

Bibliotheken – World Wide Web – professionelle Datenbanken

⇒ **Die Bibliotheken** sammeln einzelne Exemplare von Büchern, Zeitschriften und Datenträgern, die in größerer Auflage hergestellt wurden.

„**Erschließung von Literatur nach formalen Kriterien**“, also nach Arten der Themenbehandlung, z. B. Sachbuch, „schöne Literatur“, Lexikon, Lehrbuch, Loseblatt-Sammlung und viele mehr

⇒ **Archive** sammeln Originaldokumente, **Unikate**; „Urkundencharakter“, z. B. Parlamentsprotokolle, Verträge, Urteile, Geschäftsvorfälle

⇒ Die **wissenschaftliche Dokumentation** erstellt inhaltliche Erschließungen von Texten und fügt „Mehrwert“ hinzu. Dokumentationssysteme sind im Studium wichtig.

Hierzu ist der Aufbau von Datenbanken notwendig. Das Problem: man **sieht den Inhalt nicht**, sondern muss ihn mit einem **Suchsystem „ertasten“**.

Einfaches Beispiel: Datenbank JUFOBASE enthält Texte von Siegerarbeiten des Wettbewerbs „Jugend forscht“

Öffentliche Bibliotheken sind Einrichtungen der **Infrastruktur**. Sie werden aus Steuergeldern finanziert und **meist kostenlos** zur Verfügung gestellt. Dokumentationssysteme kosten immer Geld. **Das World Wide Web ist nur scheinbar kostenlos.**

Leitgedanken der Bibliotheksnutzung am Beispiel der Holzschuh-Bibliothek

Welche sind die **wichtigsten Bücherarten** in einer Bibliothek, was leisten sie **zusätzlich zu Schulbüchern** und **wie** sind sie zu **finden**?

Die **Holzschuh-Bibliothek hat Bücher zu mindestens 17 Schulfächern** und AGs, z. B.:

- allgemeine **Nachschlagewerke**, wie Lexika, Almanache und Bücher zur Allgemeinbildung, Bilderlexika
- **Fachlexika** für die Sachgebiete Literatur, Musik, Kunst und Theologie
- Texte von wichtigen Werken der Literatur
- **Interpretationen** von Werken der Literatur und der Kunst
- Reiseführer mit interessanter Information zur Geographie und Landeskunde, Kunstbände,
- Gesetzestexte - Bücher über Musik - Anleitungen zur künstlerischen Gestaltung
- „erzählende Versionen“ von Fachbüchern zur Geschichte und historische Romane, sowie
- **Biographien** (mehr als 100 vorhanden) enthalten regelmäßig interessante Informationen zur Geschichte der Lebensperiode und des persönlichen Tätigkeitsgebietes zu entnehmen,

Wichtig: Anspruchsvolle Sachbücher enthalten am Schluss eine **Bibliographie**, d. h., eine Liste der Quellen, aus der der Autor seine Information bezogen hat. Hier bekommt man Hinweise auf weitere Bücher zum Thema: Suche in **digitalen Bibliothekskatalogen!**

Zum Finden von Büchern in einer Bibliothek

Die Allgemeine **Systematik** für Bibliotheken „**ASB**“ legt die **Stellordnung** fest. Man merke sich, in welchen Abschnitten der Systematik sich welche **Schulfächer** verbergen, und umgekehrt; siehe: „**gelbes Blatt**“ und die Liste der **Übungsvorschläge**.

B. Das World Wide Web im Internet

Das Web ist eine gigantische Sammlung von Medienangeboten, die aus den unterschiedlichsten Gründen für die verschiedensten Zwecke ohne Plan geschaffen wurde. Es ist kein getreues Abbild der Wirklichkeit. Nur ein **winziger Bruchteil** dient **Bildungszwecken**. Vielfach ist das Internet/WWW nur **Datentransportmedium**, andere Angebote wurden nur für das Web geschaffen, z. B. die Systeme des Web 2.0, wie SchülerVZ, Facebook. Blogs und ähnliche.

Wir interessieren uns für den **Informationsgehalt** des **WWW** für **schulische** und **Allgemeinbildungszwecke**.

Drei Eigenschaften des WWW sind besonders wichtig:

- Der normale Nutzer kann sein **Gegenüber nicht zweifelsfrei identifizieren**.
- Jeder **Besuch im Web hinterlässt Spuren**, die von Fachleuten ausgewertet werden können.
- => Es gibt **keine Privatheit im „Netz“**: jede Aktivität kann erfasst und ausgewertet werden (z.B. mit „traceroute“). Google speichert alles auf Dauer und präzisiert damit seine Trefferlisten. Damit werden **Nutzerprofile** erzeugt, die zumindest für Werbezwecke routinemäßig ausgebeutet werden (auch im „Web 2.0“, „soziales Web“). **Zielpersonen-** statt Zielgruppen**werbung** wird möglich.
- Bis auf allererste Ansätze kann keine Suchmaschine im WWW wirklich „lesen“ (Fachbegriff „semantisches Web“), d. h. jedes Suchsystem kann nur Zeichenfolgen mit einander vergleichen.

Die Menge der Information (Anzahl der Sites) im Web ist nicht bekannt, ebenso wie der Anteil, der von Suchmaschinen erfasst wird.

Es gibt Bereiche, die nicht von Suchmaschinen erfasst werden können, weil sie kostenpflichtig sind.

Man beachte: **Bibliotheken** enthalten nur **professionell** hergestellte Produkte, die i. d. R. urheberrechtlich geschützt sind, wenn der Autor nicht schon länger als 70 Jahre verstorben ist.

Wikipedia wird **auch von Laien** gestaltet und ist (weitgehend) frei von **Urheberrechten**.

Wertvolle, **aktuelle wissenschaftliche Information** ist im Netz nur als „Abstract“ von Fachaufsätzen erhältlich. Die Zeitschriften können nur von Abonnenten im Web gelesen werden. Dies ist in der Schule allerdings nur selten ein Problem.

Es ist aber möglich, über das Netz herauszufinden, **in welcher Bibliothek sie eingesehen werden können** (z. B. über VASCODA). Die Holzschuh-Bibliothek bietet ihre Hilfe bei der Recherche an. **ernleihe** und Bestellungen von **Kopien** kosten **Geld**.

Die **typische Vorgehensweise im Web:** => **Suchen in Google - lesen in Wikipedia**

Über **Google** („**Maschine zum Gelddrucken**“) und **Wikipedia** („**Digital-intellektuelle Nachbarschaftshilfe**“) und deren Beziehung gibt es ein Informationsblatt.

Die Information im Web: Wie erkennt man, was gut ist?

- Vorschläge für Qualitätskriterien -

Unser Arbeitsziel: Wir wollen **gute** und **informativ** **Webseiten** für **schulische** und **Allgemeinbildungszwecke** finden.

- es soll **nichts kosten**

- und wir wollen „**Kostenfallen**“ **vermeiden**

Regel 1: mehrere **Dokumente** oder Web-Sites **vergleichen**, falls vorhanden

Regel 2: Unterscheidung nach **Herkunft** und **Produktionsweise:**

a.) **Originaldokumente**, z. B. Patente, gemeinfreie Buchtexte, Wetterprognosen

b.) Information **zuverlässiger Institutionen** mit **Primärliteraturcharakter** (z.B. Behörden, Verfassungsorgane; Körperschaften, die im Lichte der Öffentlichkeit stehen, z.B. anerkannte Religionsgemeinschaften),

c.) **typische Web-Informationsprodukte**, deren tatsächliche Autoren oft nicht erkennbar sind: Wikis, Foren, Blogs etc.

Fragen an ein Web-Dokument:

1. Ist das Dokument **professionell erstellt** worden?

Professionell erstellte elektronische Medien (insbesondere „e-Learning“) bieten eine **nachvollziehbare Produktionsweise**. **Autoren** lassen Rückschlüsse auf die **Sachkunde** zu. Verlagsredakteure stellen sicher, dass kompetente Autoren tätig werden.

Professionell erstellte Dokumente und Medienprodukte sind normalerweise nicht kostenlos. Ist dies dennoch der Fall, so handelt es sich oft um zeitlich oder inhaltlich begrenzte Werbeangebote, Ausnahme: Patente.

2. Werden **Quellen** angegeben? Sind sie **nachprüfbar**?

3. **Warum** wurde ein Dokument erstellt, **welche Ziele** verfolgen die **Autoren**?

Insbesondere: Ist das Dokument frei von verborgenen Absichten, wie Werbung für Weltanschauungen, Steigerung von Verkäufen oder Selbstdarstellungsbedürfnis (Weblogs)? Neutralität ist eine wichtige Anforderung an Wikipedia-Artikel, wird aber nicht immer realisiert.

4. Sind bei typischen Web-Dokumenten, wie Wikipedia-Artikeln, **Klarnamen von Autoren** zu finden, die über **Google als Fachleute identifiziert** werden können?

Gibt die **Diskussion eines Wikipedia-Eintrags** Anlass zu Zweifeln?

5. Ist ein Dokument in sich schlüssig und verständlich? Entspricht ein Text den Kriterien eines guten Aufsatzes?

6. Unbedingt das **Impressum** lesen. Dort steht, wer eine Webseite gemacht hat.

C. Einführung in die Web- Recherche

Grundregel: Je mehr man über die Suche und die Dokumente, die man sucht, **weiß**, desto besser sind die Suchergebnisse.

Gute **Suchbegriffe** findet man, wenn man sich fragt:

- welche **Dokumentarten** könnten gefunden werden und
- wie wird der gesuchte **Sachverhalt** darin **benannt**

Zunächst: Eine **Suchmaschine** ist eine Datenbank, die sehr viele Internetseiten kennt und weiß, was drinsteht. Genauer: Welche **Zeichenfolgen** an welchen Stellen stehen, in welcher Reihenfolge, **wer sie wie oft genutzt** hat und Vieles mehr.

Es gibt **zwei Methoden**, das Wissen über **die Internetseiten (Webseiten)** zu ordnen:

Das **Webverzeichnis:** Ordnet die Seiten nach einer Rangfolge der Themen nach Ober- und Unterbegriffen: Beispiel: Yahoo.
Das Prinzip wird auch bei Wikipedia verwendet.

Die **Volltextdatenbank:** Bekanntestes Beispiel: Google

Meist werden Mischungen beider Systeme verwendet.

Neben den großen, allgemein bekannten Suchsystemen, wie Google und Yahoo, gibt es eine große Zahl von **Spezialsuchmaschinen**, die in ihrem **Fachgebiet** wesentlich bessere Treffer geben, als die Giganten des Marktes. Sie haben dafür andere Schwachstellen.

Grundsätzliches zur Suche bei Google

Die Suchergebnisse bei Google werden beherrscht von einem mathematischen Verfahren, das aus einer riesigen Zahl von Dokumenten, in denen der Suchbegriff oft enthalten ist, die wichtigsten an die vorderen Plätze rückt. Welche das sind, regiert der „**PageRank-Algorithmus**“. (Algorithmus = Arbeits- oder **Rechenvorschrift**). Die Arbeitsweise ist das wichtigste Betriebsgeheimnis einer der bestverdienenden Firmen der Welt. Jede Trefferliste einer Suche mit logischen Verknüpfungen wird vom PageRank-Algorithmus erneut geordnet.

Google bewertet einige **Dokumenteigenschaften** als sehr **wichtig** für die Bedeutung eines Dokumentes, darunter sind:

- die **Verlinkung** mit möglichst vielen anderen Dokumenten (Querverweise **von** anderen Dokumenten); diese Tatsache wird „Weboptimierern“ missbraucht, indem man von „Linkfarmen“ eine riesige Zahl von nutzlosen Links schaltet;
- eine lange **Liste von zitierten**, oder **zu** weiterführenden **Dokumenten**; auch diese Eigenschaft wird manchmal mit einer grotesk langen Zitatenliste missbraucht, um Dokumente auf vordere Positionen zu mogeln;
- die „**Metatags**“ (Stichworte zum Inhalt, von den Autoren der Webseite vergeben)
- die **thematische Nähe zu bezahlten Einträgen**, den „**sponsored Links**“
- ein Eintrag in **Wikipedia** (führt fast immer zu einer Spitzenposition in der Liste)

- Beiträge aus **Diskussionsforen**, besonders bei aktuellen technischen Themen
- **Wichtig:** Die **Zahl der Aufrufe eines Dokumentes** wird gespeichert und beeinflusst den **Sortierang**. Bei einem noch nie oder selten genutzten Dokument wirkt dies sehr stark. Auch die Suchübungen einer ganzen Schulklasse verändern die Wichtung.
- Außerdem wird das gesamte **Verhalten der Nutzer** gespeichert und ausgewertet, wie Lesedauer, Reihenfolge und Vieles mehr. Die Information über das **Leseverhalten der Webnutzer** der reichen Länder gilt als der wichtigste Wissensschatz der Suchmaschinenbranche.

A. Die Einfachversion der Google-Suche

Wie sucht man bei Google mit einem oder mehreren Begriffen in der Standard-Eingabemaske?

Ein Tipp: Macht euch klar, welche Voreinstellungen Google automatisch anwendet. Der folgende Text nennt nur die wichtigsten.

1. Wird **nur ein Begriff** zur Suche verwendet: Der PageRank-Algorithmus findet die Trefferliste

2. Gibt man **mehr als einen Begriff** ein, unterstellt Google, dass man Dokumente wünscht, die **alle** diese Begriffe **irgendwo** enthalten. **Das Ergebnis ist von der Reihenfolge der Suchbegriffe abhängig!**

Sie sind mit der **UND-Bedingung** verknüpft (Fachbegriff: **UND-Logik**).

Frei formuliert: Finde **alle Dokumente**, die **alle** eingegebenen **Worte („Zeichenfolgen“)** **irgendwo** im Dokument **enthalten**.

Groß- und Kleinschreibung wird ignoriert. **Häufige kurze Wörter** ohne Aussagekraft, einzelne Zahlen und Buchstaben, werden **nicht berücksichtigt** (können aber hinzu genommen werden).

Google sucht **automatisch nach möglichen Schreibfehlern** und bietet die - vom Programm automatisch „vermutete“ - richtige Variante an, oder sucht alle bekannten Buchstabiervarianten automatisch. Dies vereinfacht die Suche nicht immer.

Schlussfolgerung: Die Nutzerfreundlichkeit von Google führt manchmal zu Schwierigkeiten. Man mache sich die Voreinstellungen bewusst!

B. Die Fortgeschrittenenversion der Google-Suche: „erweiterte Suche“

Die folgende Darstellung folgt den „Grundlagen der Google Suche“, die wie folgt zu erreichen sind:

Google => erweiterte Suche => Suchtipps => Websuche Hilfe => Such-Grundlagen

Für die folgenden Überlegungen sind **Grundkenntnisse** der **Mengenlehre** nützlich.

Es geht um die Suche bei Google mit **mehreren Begriffen**, die mit verschiedenen „**logischen Verknüpfungen**“, in den Treffern auftauchen können.

Es kann gesucht werden nach **Dokumenten**,

die **mehrere Begriffe** gleichzeitig enthalten,

die **eine bestimmte Wortfolge** enthalten,

die **mindestens einen von mehreren** Begriffen enthalten,

die bestimmte **Begriffe nicht enthalten**.

Diese Suche kann eingegrenzt werden, auf den **Titel**, den **Textkörper**, die **URL**, oder auf die **Links**.

Hierzu kann man die **Eingabemasken der „erweiterten Suche“** verwenden.

1. Suche von Dokumenten, in denen alle gewünschten Suchbegriffe auftauchen.

Man spricht von einer Verknüpfung mit „**Und-Logik**“. Grundsätzlich können so Antwortlisten erhalten werden, die eher viele zutreffende Dokumente enthalten und wenig „Ballast“. Es ist jedoch nicht ausgeschlossen, dass wichtige Treffer ausgeschlossen wurden, weil sie nicht alle verlangten Begriffe enthalten haben.

Wichtig: **Prüfen, welche Begriffe sind zwingend** wichtig, und welche verkleinern den Antwortsatz unnötig!
Auch an **Synonyme** denken, das macht Google manchmal, aber noch **nicht immer!** Das ist ein Thema der **semantischen Suche**, dem Web der Zukunft, an dem gearbeitet wird.
Eingabereihenfolge beeinflusst das Ergebnis.

2. Suche von Dokumenten, in denen mehrere Begriffe in exakter Reihenfolge enthalten sind, z.B. vollständige **Namen, ganze Sätze, oder Satzteile**.

Fachbegriff: „**Wortgruppensuche**“, auch „Bound - Phrase-Suche“

Alternative: Statt der Eingabe in das vorgegebene Feld der „erweiterten Suche“, genügt es auch, die Wortfolge im **normalen Eingabefenster in Anführungszeichen** ein zu schließen.

Hiermit können Texte der Literatur mit der Eingabe eines Titels oder eines Teils eines Satzes gefunden werden.

3. Suche von Dokumenten, in denen **mindestens einer von mehreren Begriffen** auftaucht.

Fachbegriff: „**Oder-Logik**“

genauer: das „**einschließende oder**“

4. **Suche von Dokumenten, die, neben den gewünschten, nach Punkt 1 bis 3, einen oder mehrere Begriffe nicht enthalten, um den Antwortsatz zu bereinigen.**

Fachbegriff: „**Not-Logik**“

Doppeldeutigkeit mancher Suchbegriffe ist ein häufiger Grund hierfür.

Man sollte hiervon nur **selten Gebrauch** machen, weil nicht absehbar ist, ob ein solcher Begriff durch Zufall in einem interessanten Dokument enthalten ist. Es würde ausgeschlossen werden, ohne dass wir es je bemerken. Hinzunahme von mehr Suchbegriffen in UND-Verknüpfung ist besser zur Präzisierung geeignet.

Zusammenfassung der Suchstrategie im World Wide Web:

Diese Überlegungen gelten, wenn **mehr als** das **Wikipedia**-Dokument gefunden werden soll.

1. Überlegt euch, wie euer **ideales Dokument** aussieht.
2. Fragt euch, wie das, was ihr sucht, **in diesem Dokument benannt** sein kann.
3. Die Suche mit **allgemeinen Begriffen** beginnen (oder sogar in ODER-Logik-Verknüpfung), um sicher zu sein, dass man überhaupt etwas finden kann.
4. Durch **Lesen der ersten Trefferdokumente** findet man oft genauere Begriffe, die wiederum in der nächsten Fragestufe eingesetzt werden können.
5. Findet man viele Treffer, kann man mit engeren Begriffen oder mit mehreren in einer **Anfrage** die Menge **präzisieren** (Verknüpfung mit „UND“-Logik).
6. **Dokumente**, die man durch eine sofortige, **scharfe Auswahl** gar **nicht erst sieht**, sind **sofort verloren**.
7. Merken, in welchen **Web-Dokumenten** man zu welchen **Fragen gute Information** gefunden hat.
8. Aus den **eigenen Erfahrungen**, langfristig eine **Hitliste** aufbauen, wo man zu welchen Fächern – über Wikipedia hinaus – gute **schultaugliche Information** herbekommt.