



# Internet und die Folgen

Wissenschaftlicher Bibliothekarenkurs  
2001/2002

© 2002 Trialog AG

# Lernziele

Sie wissen,

- ☛ wie Sie Informationen im Internet beschaffen.
- ☛ wie Sie selbst Informationen im Internet publizieren.
- ☛ Wie das Internet funktioniert.
- ☛ Wer Informationen anbietet.
- ☛ ...

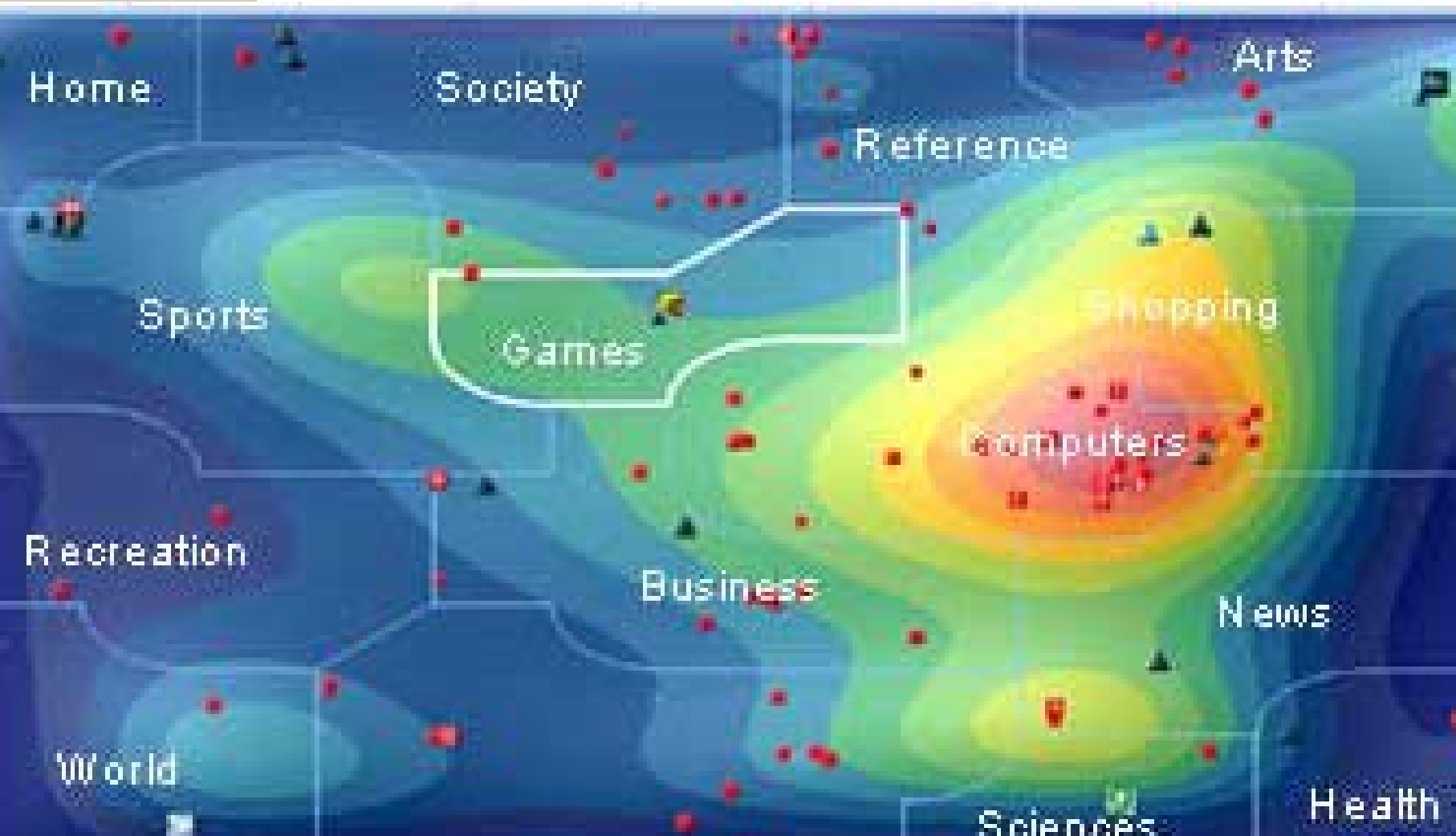
# Programm 23.05.2002

- ☛ Einleitung
- ☛ Was wollen Sie wissen?
- ☛ Wie gross ist das Internet?
- ☛ Suchwerkzeuge
- ☛ Suchstrategien

# Grösse des Internet

- ✂ 1997: ca. 800 mio Seiten
- ✂ Mitte 2000: ca. 2 mia Seiten
- ✂ Heute: ca. 3 mia Seiten
  
- ✂ Schätzung: Es ist nur 1/500 erreichbar

# Thematik des Internets





# Was findet sich nicht im Internet?

🔧 (interaktiv)

A decorative green line starts from the top left, curves across the top, and ends with a small green sphere on the left side. The sphere has a black outline and contains several small white dots.

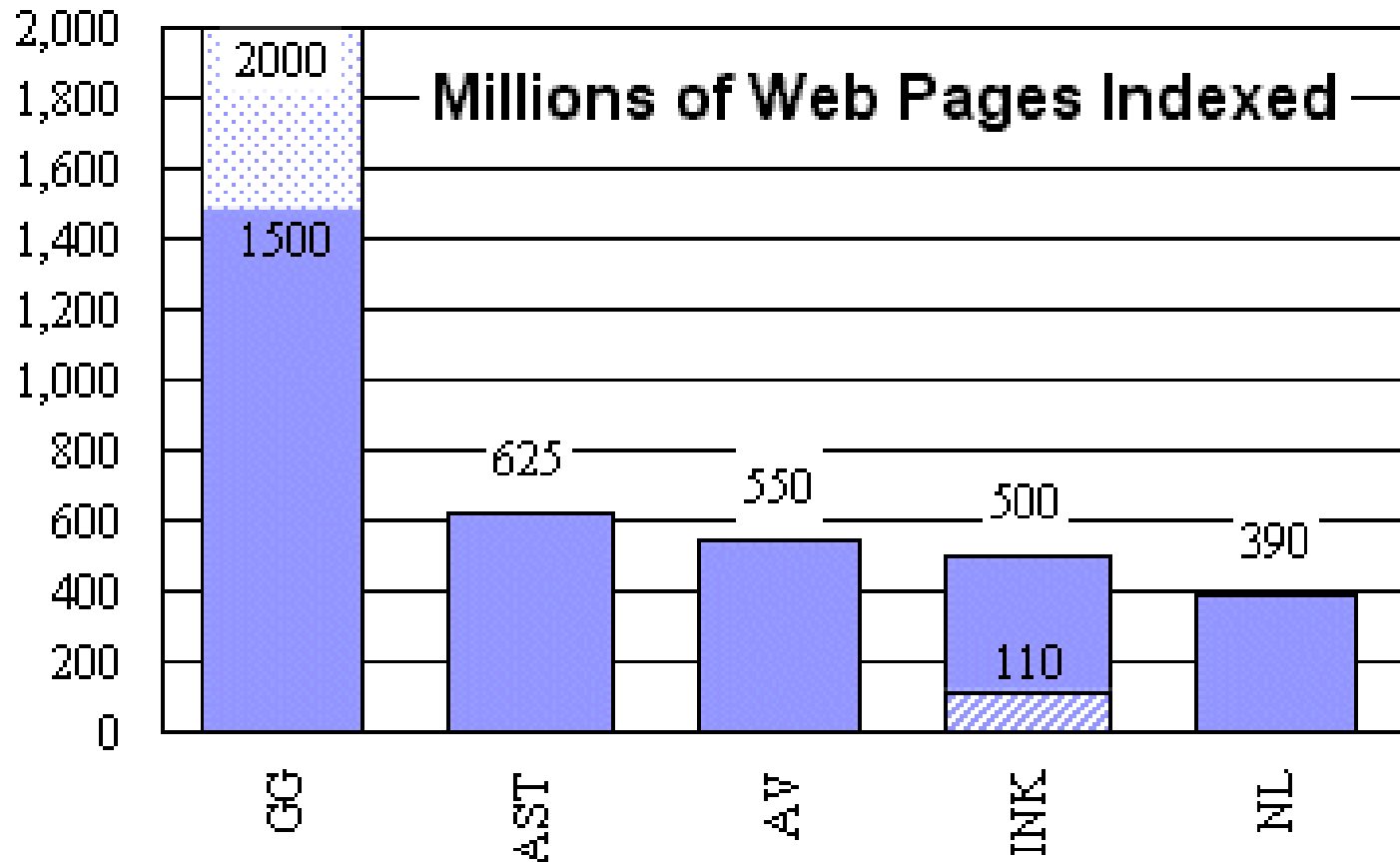
# Suchmaschinen

# Namen

- 🐞 **Fast**
- 🐞 **Northern Light**
- 🐞 **AltaVista**
- 🐞 **Google!**
- 🐞 **Excite**
- 🐞 **Infoseek**
- 🐞 **Lycos**
- 🐞 **WebCrawler**
- 🐞 **Inktomi:** MSN Search / iWon / Yahoo! / AOL / Anzwers / HotBot



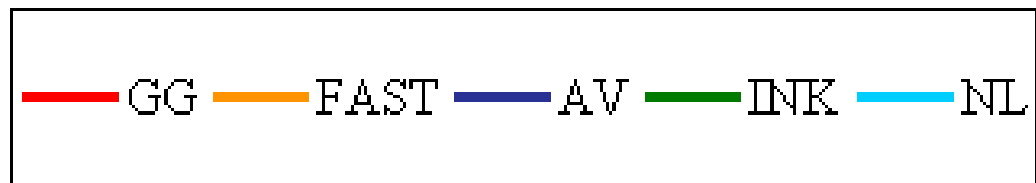
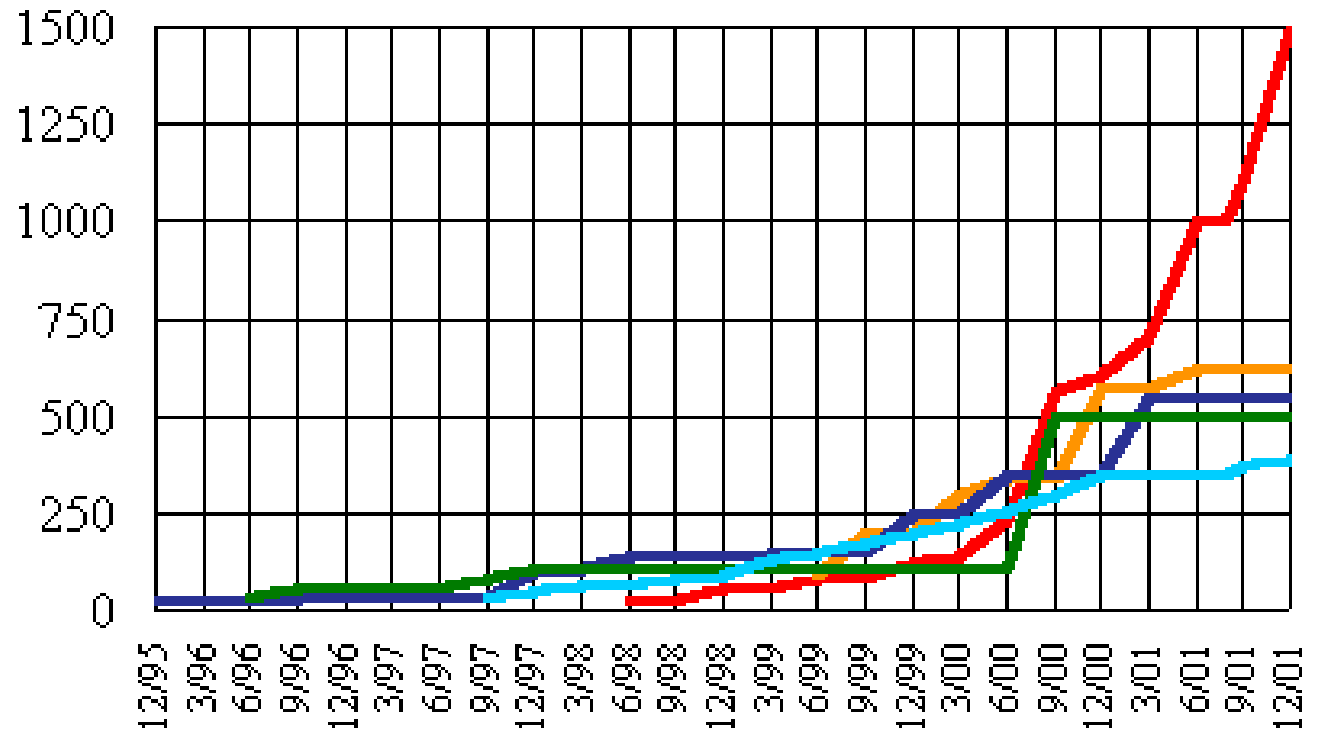
# Grösse von Suchmaschinen



# Entwicklung von Suchmaschinen

## Search Engine Sizes

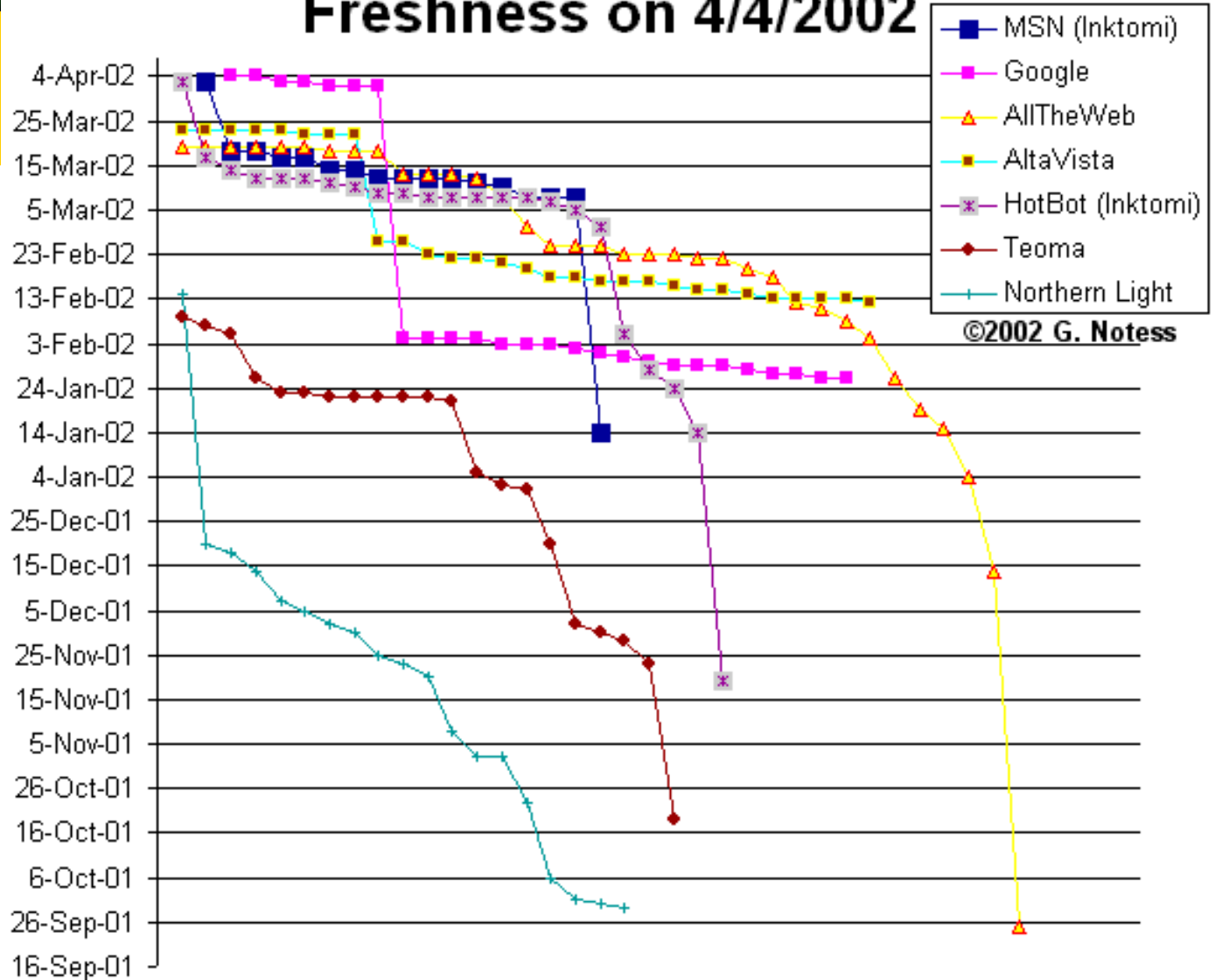
(millions of web pages)



# Aktualität (1)

Search Engine	Neueste Seite	Älteste Seite
Google	1 Tag	68 Tage
MSN (Inktomi)	1 Tag	80 Tage
HotBot (Inktomi)	1 Tag	136 Tage
AltaVista	12 Tage	51 Tage
AllTheWeb	16 Tage	191 Tage
Teoma	54 Tage	167 Tage
NLResearch	49 Tage	187 Tage
WiseNut	247 Tage	286 Tage

# Freshness on 4/4/2002

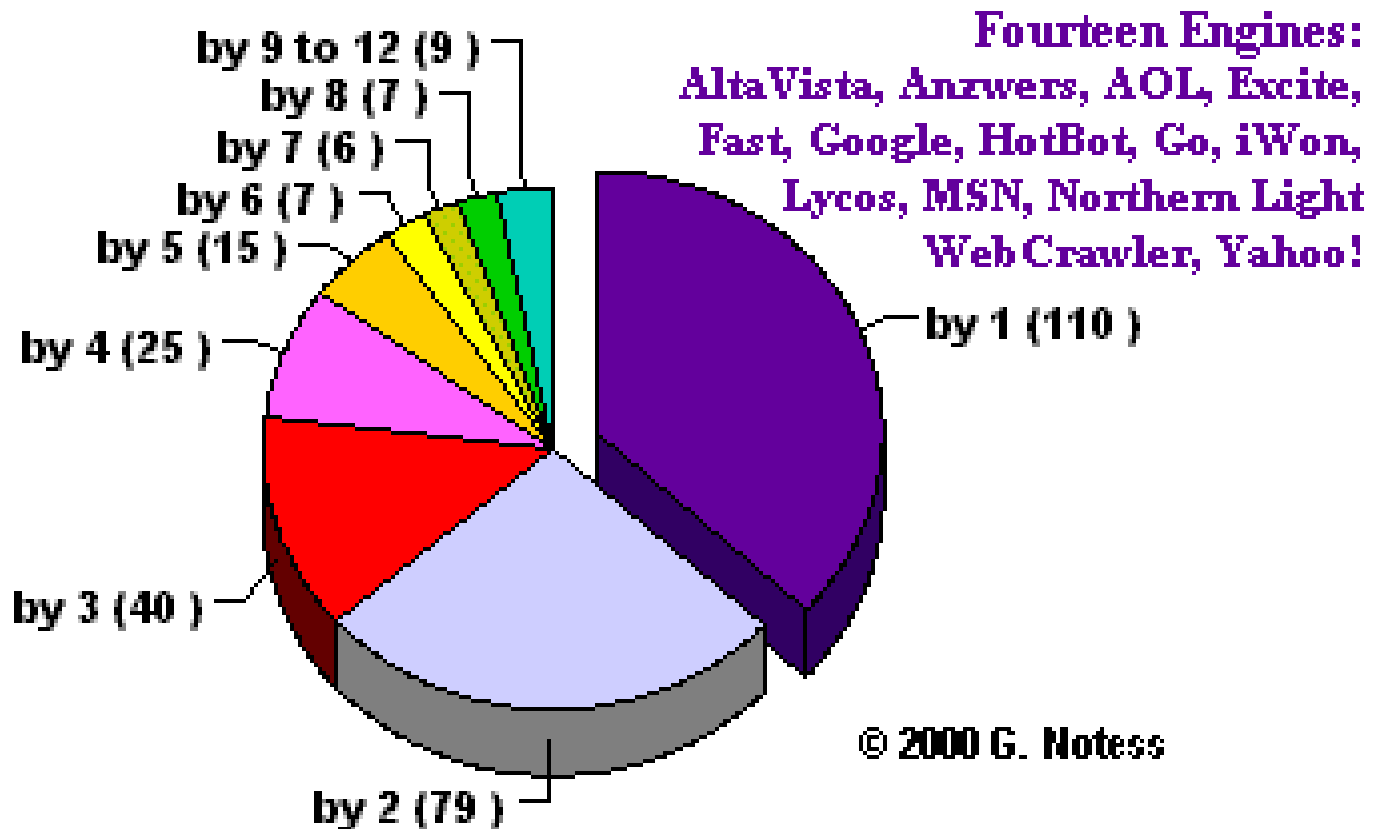


©2002 G. Notess

# Überschneidung

## Overlap of Five Searches

298 Total Unique Hits on 2/21/00



# Beispiel Google

- ☛ Ist sehr umfangreich.
- ☛ bewertet die Webseiten als wichtig, auf die viele andere Webseiten verweisen (top!)
- ☛ „Ähnliche Seite,,
- ☛

# Suchen im Internet

🔦 „Wer sucht, der findet.“

Frage nur: Wann?

🔦 „Information auf Knopfdruck“

Wie lange dauert es, bis Sie den entscheidenden Knopfdruck machen können?

# Zugang zu Informationen

## Mit folgende Methoden:

- ✚ URL
- ✚ Navigieren („Browsen“)
- ✚ Suchen
- ✚ Erhalten

## Art der Lieferung:

- ✚ „Push“ (E-Mail, LISTSERV) or „Pull“ (Web, FTP, News)



# Wie funktionieren Suchmaschinen?



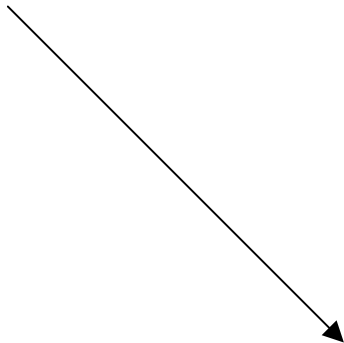
Suchanfrage



Sammeln von  
Dokumenten

Indexieren

Abgleichen  
von Anfrage  
und Dokumente



# Was indexieren Suchmaschinen?

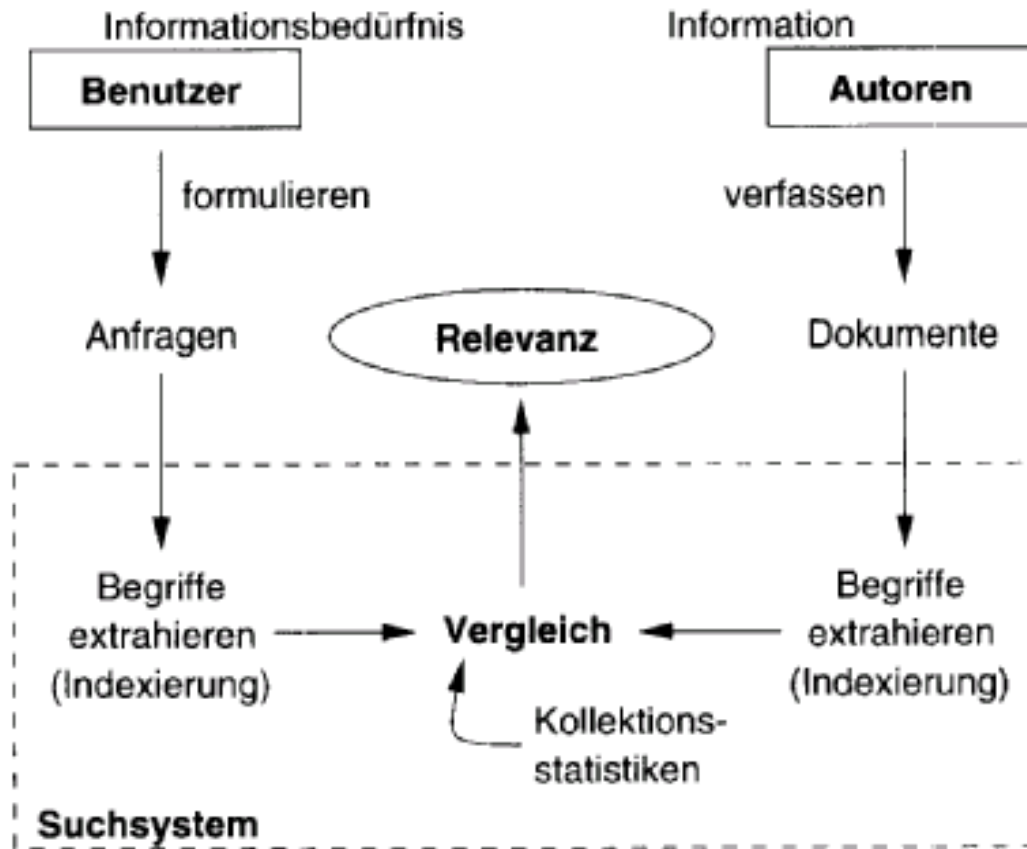
## **Frei zugänglich:**

- ☛ „Öffentlicher“ Webserver (HTML, PDF, WORD, ...)
- ☛ Datenbank (über Webserver erreichbar)

## **Nur eingeschränkt zugänglich:**

- ☛ Kostenpflichtige Server
- ☛ Intranet

# Abgleich Anfrage $\Leftrightarrow$ Kollektion



# Relevanz

- 🔦 Geschätzte Relevanz  
(„Was die Suchmaschine anzeigt.“)
- 🔦 Subjektive Relevanz  
(„Was mir etwas gebracht hat.“)
- 🔦 Objektive Relevanz  
(„Was sachlich zutrifft.“ Experte)

# Rangierprinzipien

- ✚ Je **mehr Suchbegriffe** in einem Dokument vorkommen, desto wahrscheinlicher ist das Dokument relevant.
- ✚ Je **häufiger ein Suchbegriff** in einem Dokument vorkommt, desto wahrscheinlicher ist das Dokument relevant.
- ✚ Dokumente, die **seltene Suchbegriffe** enthalten, sind mit einer höheren Wahrscheinlichkeit relevant als Dokumente, die häufige Suchbegriffe enthalten.

# „Pitfalls“

Tipps:

- 🚫 Suche: klein, GROSS oder gemischt?
- 🚫 Wurden die Infos erfasst?
- 🚫 Sind Dokumente für Roboter gesperrt?

# Suchen im Internet (Quellen)

✎ <http://wwwplus.pctip.ch/finden/>

✎ [www.suchfibel.de](http://www.suchfibel.de)

✎ [www.internet-kompetenz.ch](http://www.internet-kompetenz.ch)

✎ <http://www.fbi.fh-koeln.de/service/internet/internet.htm>

## **Papier:**

✎ Werner Hartmann: Informationsbeschaffung im Internet.  
Grundlegende Konzepte verstehen und umsetzen.

✎ Gale guide to Internet databases

# Suchmaschinen & Katalog

- 🐞 [www.google.com](http://www.google.com)
- 🐞 [www.alavista.com](http://www.alavista.com)
- 🐞 [www.hotbot.com](http://www.hotbot.com)
- 🐞 [www.infoseek.com](http://www.infoseek.com)
- 🐞 [www.excite.com](http://www.excite.com)

Deutsch:

- 🐞 [www.altavista.de](http://www.altavista.de)
- 🐞 [www.fireball.de](http://www.fireball.de)

- [www.yahoo.com](http://www.yahoo.com)
- [www.snap.com](http://www.snap.com)
- [www.looksmart.com](http://www.looksmart.com)
- [www.newhoo.com](http://www.newhoo.com)
- ...

Deutsch:

- [www.yahoo.de](http://www.yahoo.de)
- [web.de](http://web.de)



# Bewertung von Suchdiensten

## STIFTUNG WARENTEST: DREI MIT NOTE GUT

- ☛ Die Stiftung Warentest nahm mit konkreten Suchaufträgen 14 Suchmaschinen unter die Lupe. Dabei wurden Altavista, Fireball und Google als gut benotet.
- ☛ <http://www.altavista.de/>
- ☛ <http://www.fireball.de/>
- ☛ <http://www.google.com/>

Quelle: Search & Find

# Ausgangspunkte

- 🔗 [www.suchfibel.de](http://www.suchfibel.de)
- 🔗 [www.info-search.de/suche/](http://www.info-search.de/suche/)
- 🔗 [www.ch/](http://www.ch/)
- 🔗 <http://www.suchmaschinenindex.de/>
- 🔗 <http://www.ertools.ch/>
- 🔗 ...
- 🔗 [www.download.com](http://www.download.com)



# Tipps für Suchstrategie (1)

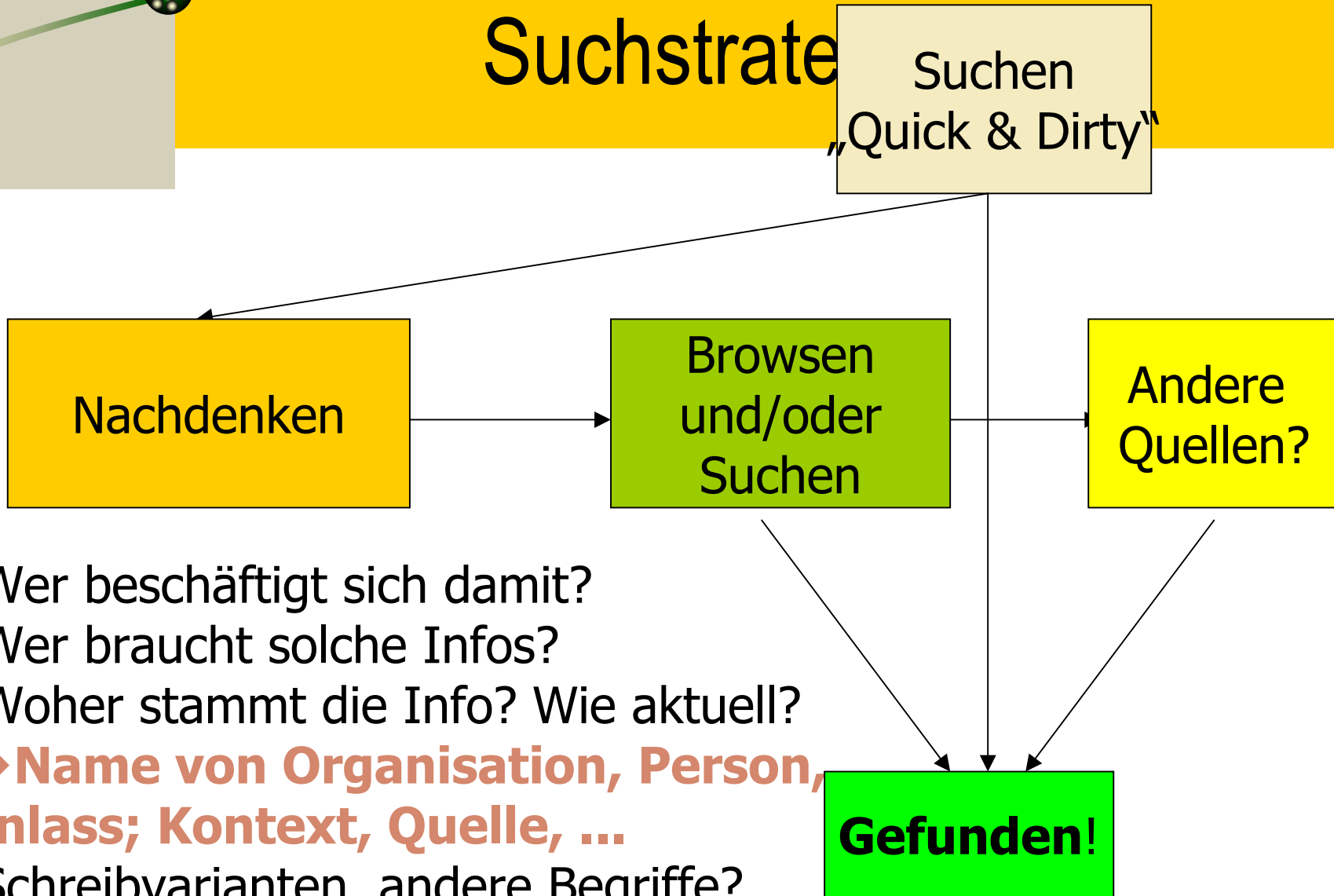
1. Richtige Dokumentensammlung wählen
2. Richtiges Werkzeug benutzen
3. Viele präzise Suchbegriffe verwenden
4. Unzutreffende Dokumente in Trefferlist ignorieren
5. Das eigene Wissen in Suchbegriffen einbringen



# Tipps für Suchstrategie (2)

6. Erste Treffer-Dokument(e) vollständig untersuchen
7. Interaktive Techniken anwenden
8. In Teilkollektionen suchen
9. Anfragen in verschiedenen Sprachen formulieren
10. Rechtschreibung and alternative Schreibweisen beachten

# Suchstrategie



- Wer beschäftigt sich damit?
- Wer braucht solche Infos?
- Woher stammt die Info? Wie aktuell?
- **Name von Organisation, Person, Anlass; Kontext, Quelle, ...**
- Schreibvarianten, andere Begriffe?
- Anderer Weg** (Papier, Telefon, ...)

**Gefunden!**

# Bewertung von Information

## **Recherchebezogen:**

- ✂ Relevant?
- ✂ Vollständig?
- ✂ Aktuell?

## **Quellenbezogen:**

- ✂ Verlässlich? Authentisch?
- ✂ Verfügbar? (elektronisch, erschlossen)

## **Nutzungsbezogen:**

- ✂ Lesbar? Gut dargestellt? Weiter bearbeitbar?  
Speicherbar?

# Informationsquellen

- 🔗 <http://www.searchenginewatch.com/>
- 🔗 <http://www.searchengineshowdown.com/>
- 🔗 <http://www.webmap.com/trademapdemo.html>
- 🔗 <http://www.vibrio.de/service/altavist/texte/090200.pdf>
- 🔗 <http://www.invisibleweb.com>

A decorative green line starts from the top left, curves down and right, passing through a small sphere with a grid pattern, and then continues across the top of the slide. The sphere is positioned at the top left corner of the yellow bar.

\*\*\* ENDE \*\*\*



# Die verschiedenen Formate

## **Daten:**

- Text: \*.txt, \*.doc, \*.wp, \*.htm; \*.pdf
- Bild: \*.gif, \*.jpg; \*.bmp, \*.png, \*.tif
- Audio: \*.au; \*.wav; \*.mp3
- Video: \*.mpg, \*.mpeg
- Spezial: \*.dat, \*.tab, \*.db, \*.mdb, \*.xls

## **„Ausführbar“: \*.exe, \*.com**

- Programme (Software)

## **Archive: \*.zip, \*.gzip, \*.tar, \*.sea**

- Zusammenfassung von Dateien (mit Pfad)

## **Komprimiert: \*.zip, \*.tgz, \*.hqx**

- Gekürzt + Fehlerkorrektur

# Häufige Datenformate

.htm, .html	Hypertext	Browser
.txt	ASCII-Text	Browser, ...
.jpg, .gif	Bildformate	Browser, Bildbearbeitungsprogramme
.zip	Komprimierte Datei	Unzip, Winzip
.doc	Textverarbeitung	Word
.pdf	„Digitales Papier“	Acrobat

# Bewertung von Formaten

X = ja    O = eingeschränkt    - = nein

Anforderung an das Dateiformat / Dateiformat	MS Word	ASCII-TeXt	PS	PDF	RTF	TeX,	SGML	HTML	XML <sup>(4)</sup>
erfügb. Erstellungswerkzeuge	O	X	X	X	X	X	O	X	X
erf. Bearbeitungswerkzeuge	O	X	-	-	X	X	O	X	O
strukturierbarkeit	O	-	-	-	O	O	X	O	X
recherchierbarkeit	O	O	O	O	O	O	X	O	X
konvertierbarkeit	O	X	-	-	X	X	X	X	X
präsentation	O	X	X	X	O	X	O	X	O
hypertextfähigkeit	O	-	-	X	O	O	X	X	X
portierbarkeit	X	-	X	X	X	O	O	O	O
standardisierung	-	X	O	O	O	O	X	X	X
archivierbarkeit	-	X	O	O	O	O	X	O	X

# Software für Formate (Kompression)

- ✂ \* .zip WinZip / pkunzip (Windows)
- ✂ \* .Z uncompress oder "gzip -d,,
- ✂ \* .tar tar (UNIX-Backup-Programm)
- ✂ \* .tgz zuerst gunzip, dann tar
- ✂ \* .tar.Z zuerst uncompress, dann tar
- ✂ \* .hqx binHex verwenden (Mac)

# Wenn das Format nicht stimmt

🚫 Probleme Mac ↔ PC ↔ Unix

– Falscher Zeichensatz

🚫 Falsche Programmversion

🚫 Fremde Zeichensätze (Unicode)

⇒ In Notepad anschauen. / Rückfragen / anderes  
Format verlangen / Konversion /  
Spezialistenarbeit

# Zeichensätze sind verschieden

Windows	Mac	DOS (CP850)	UNIX
schöne Mélisande	sch <sup>^</sup> ne MÈlisande	sch÷ne MÚlisande	sch÷ne MÚlisande
246 → ö 233 → é	246 → ^ 233 → È	246 → ÷ 233 → Ú	246 → ÷ 233 → Ú

# Recherchen

## 🔍 Recherchen im Internet (Bsp.)

1. Was ist die Zukunft der wiss. BibliothekarInnen?
2. Welche bibliothekarischen Stellen sind in Köln offen?
3. Was bedeutete die Allianz von Bertelsmann und Napster?