

Google

Reginald Ferber

Hochschule Darmstadt

Fachbereich Media, Studienbereich Informationswissenschaft

reginald.ferber@h-da.de

Google

Plan:

Rückblick + Entwicklung Web / Suchmaschinen

Idee + Technische Grundlagen

Finanzierung + Interessenlagen

Ansätze + Entwicklungsperspektiven

Geschichte

2009:

"Wert der Marke Google" > 100 Mrd \$ (seit 2007 "wertvollste Marke")

1999:

Das World Wide Web (Web) wird zu einem Alltagsmedium (Google wird ein Jahr alt)

1989:

Das Web existiert noch nicht (das Internet schon).

1979:

Das Internet verbindet vor allem Rechenzentren an Hochschulen und Forschungseinrichtungen. E-Mail beginnt.

1969:

Das ARPA-Net, der Vorgänger des Internet, wird eingeführt.

Geschichte des Web / Suchmaschinen

1989:

Erster Entwurf für das WorldWideWeb (WWW) von T. Berners-Lee am CERN

1991/92:

Erste Versionen und Server am CERN und darüber hinaus

1994:

Erste internationale WWW-Konferenz (Genf)

Yahoo! startet (unter anderem Namen) als Web-Verzeichnis

Erste Web-Suchmaschinen (WebCrawler, Lycos, Excite)

1995/96:

AltaVista von Digital Equipment Corporation (DEC) bekannteste SM

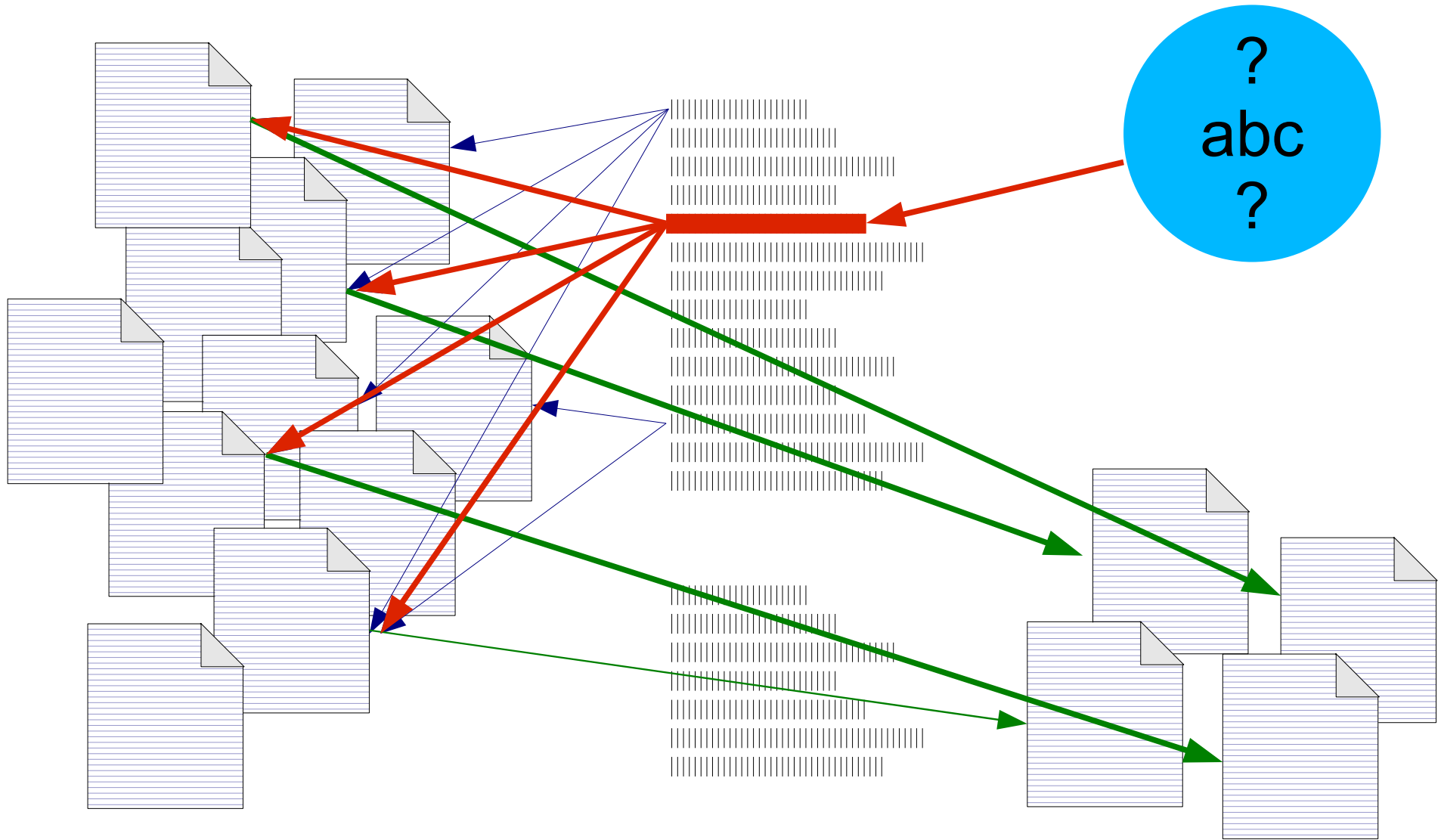
1998:

Google startet mit einem neuen Suchmodell (PageRank)

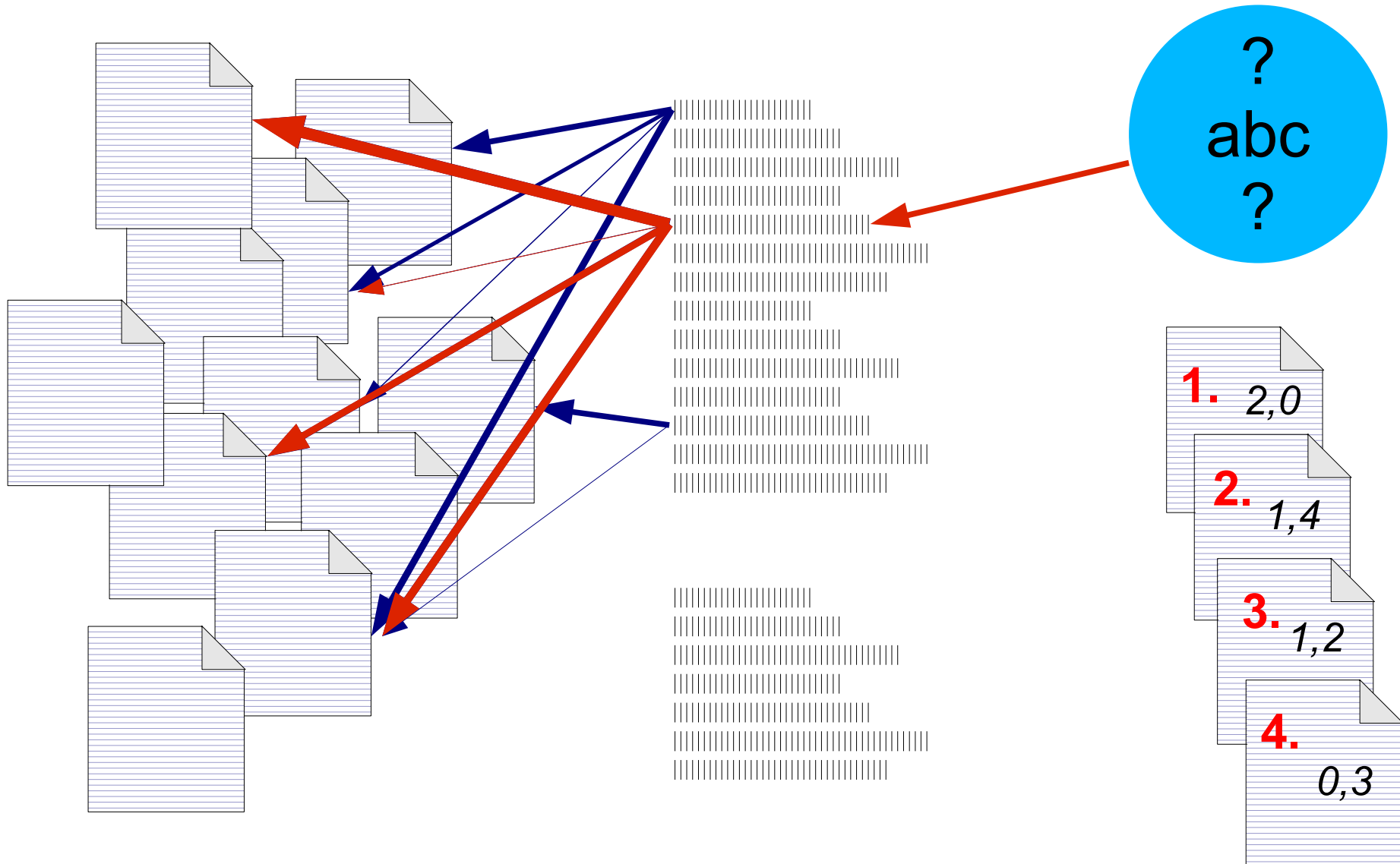
2000:

Google wird zur am meisten genutzten Suchmaschine weltweit.

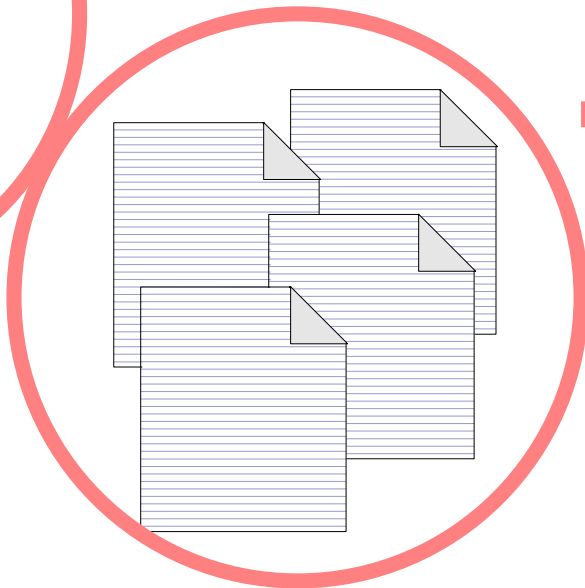
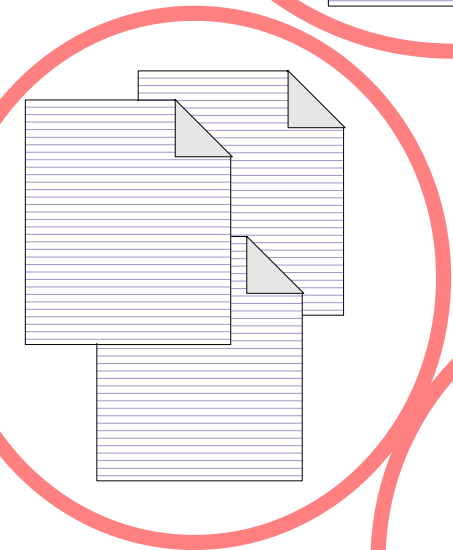
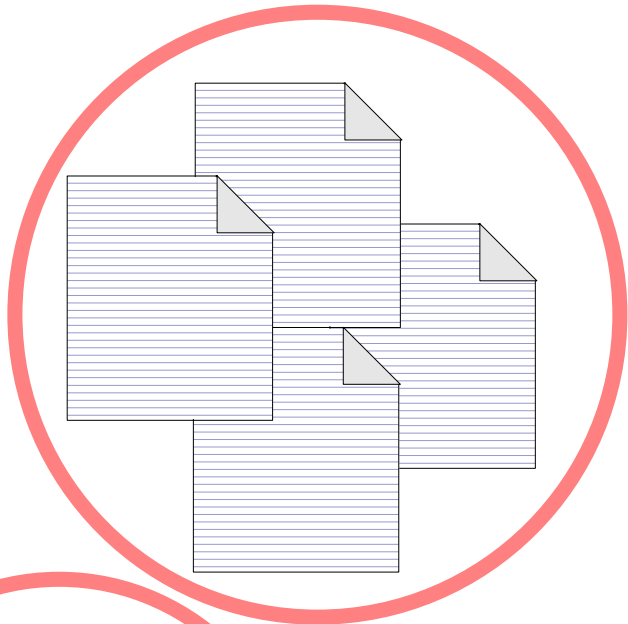
Suchmaschine



Ranking



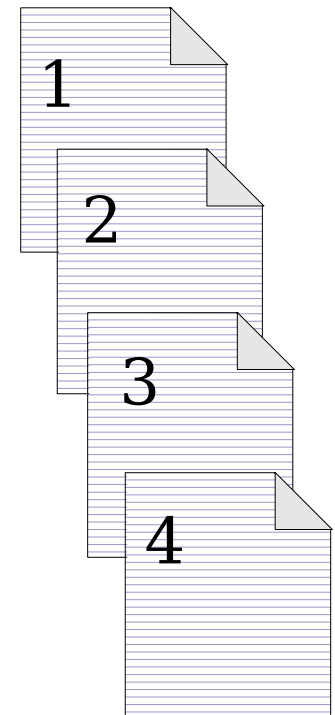
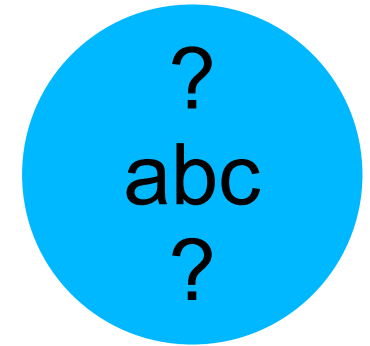
Websuche











Eigenschaften Websuche

Wachstum des Web:

- Zahl der Server und Seiten wird immer größer
- Web-Suchmaschinen werden "Gatekeeper" zum Netz

Heterogenität:

Im Web finden sich sehr viele unterschiedliche

- Themen
- Anbieter
- Nutzende
- Dienste
- Dokumentformate
- Sprachen

Screen-Design statt Informations-Design

- Es gibt wenig zentrale Vorgaben
- Inhalt vieler Seiten lässt sich schlecht automatisch erfassen

Kommerzielle Interessen:

- Konkurrenz zwischen Anbietern gleicher oder ähnlicher Dienste um die ersten Plätze in den Ergebnislisten
- Optimierung der Seiten nach den Gewichtungsmethoden der Suchmaschinen / Spamming

--> Es wird immer schwieriger / aufwändiger gute Suchergebnisse zu erzielen

Google Start

Neues Konzept:

- Ranking hängt nicht (nur) von der einzelnen Seite ab, sondern von ihrer Position im Web

genauer: von den Verweisen, die auf sie zeigen.

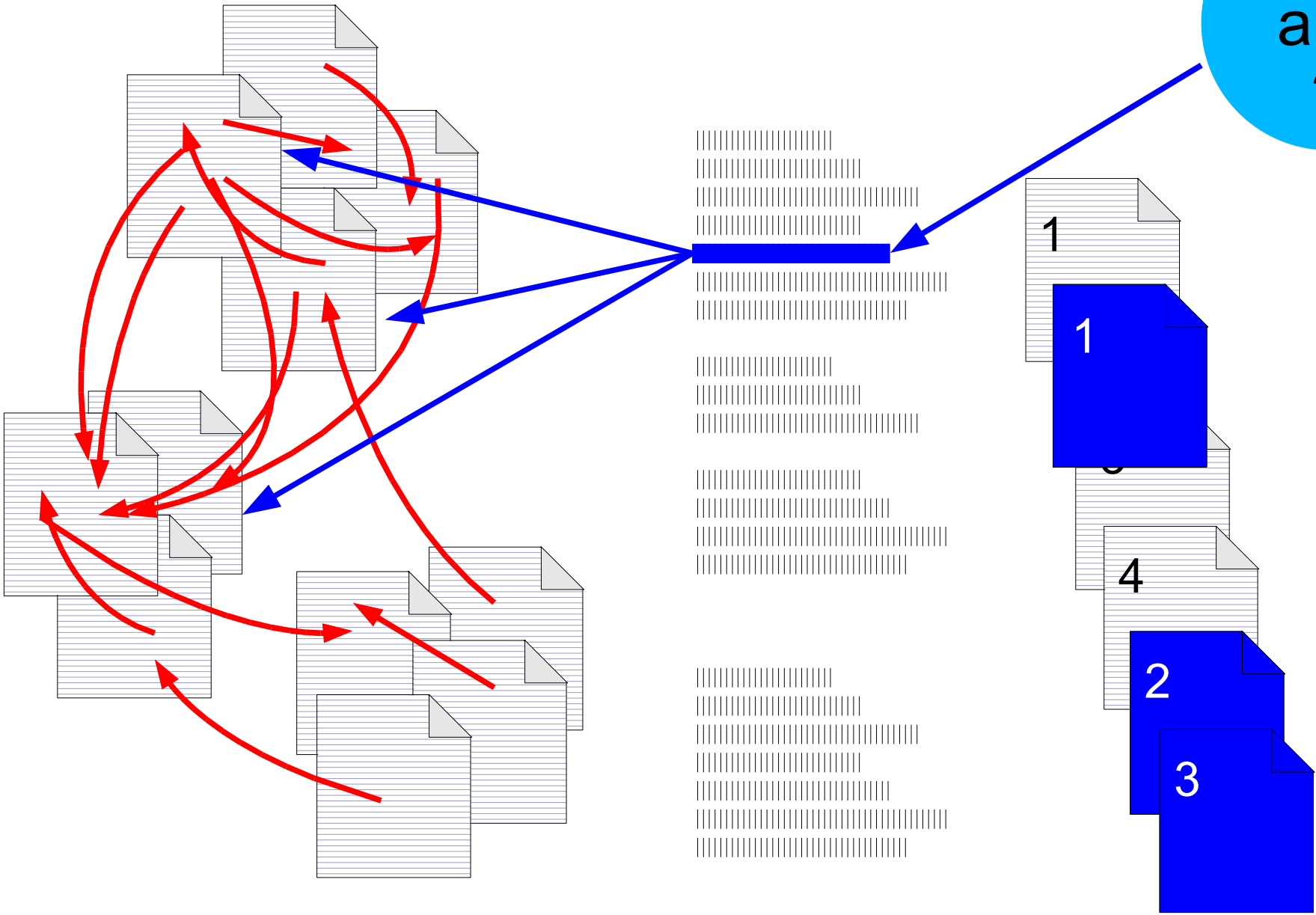
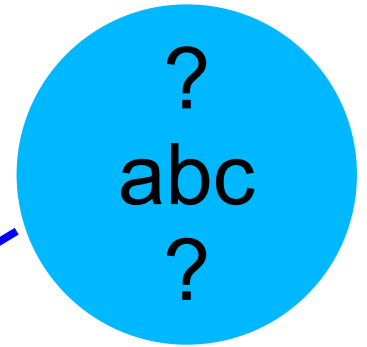
PageRank[®]-Algorithmus: Seiten, auf die oft (von guten Seiten aus) verwiesen wird, sind gute Seiten und werden hoch gerankt.

- Globale Reihenfolge aller Seiten
- Ergebnismenge zu einer Anfrage, wird nach dieser globalen Reihenfolge sortiert.

Zusätzlich:

- sehr schneller Algorithmus (auf Clustern von Linux Rechnern)
- gute Skalierbarkeit, hohe Ausfallsicherheit
- sehr einfaches, eingängiges Layout
- sehr viele Seiten erfasst (großer Index)
- gute PR

PageRank



Finanzierungsmodelle

Klassische Ansätze:

- **Nutzende** bezahlen
(Abo mit Login, Modell klassischer Datenbankdienstleister)
- **Seitenanbieter** zahlen für die Aufnahme
- **Werbung** auf Such- und Ergebnisseiten

Alle drei Modelle haben Probleme.

Suchmaschinenspezifisches Modell:

- Seitenanbieter zahlen dafür, dass ihre Anzeige (mit Link auf die eigene Seite) **bei bestimmten Suchtermen** eingeblendet wird.
- Erste Versuche 1996 scheitern (Open Text).
- GoTo.com (später Overture, dann Yahoo!) ist 1998 erfolgreich.
- Google bietet "AdWords" ab 2000.

Variante:

Bezahlen, wenn der Link angeklickt wird (pay per click)

Interessenlagen

Das Finanzierungsmodell führt zu spezifischen Interessenlagen:

Nutzende:

- Möglichst gute / objektive (?) Ergebnisse (vollständig, beste zuerst)
- Einfache Bedienung, schnell, geringe Kosten

Kommerzielle Seitenanbieter:

Bei relevanten (?) Suchtermen auf den Ergebnisseiten gut sichtbar sein

- weit oben im Ranking (der Ergebnisliste = "organic results")
- Anzeigen ("sponsored results")
- viele Suchende mit möglichst spezifischem Interesse

Suchmaschinenanbieter:

- hohe Einnahmen aus Stichwortverkauf
 - viele "AdWord"-Kunden
 - hohe Preise für Stichworte
 - viele Anfragen mit verkauften Suchtermen
 - viele Suchende
 - für Suchende attraktive Ergebnisse
 - Anzeigen den "richtigen" Suchenden zeigen

Möglichkeiten, die SM-Ziele zu erreichen

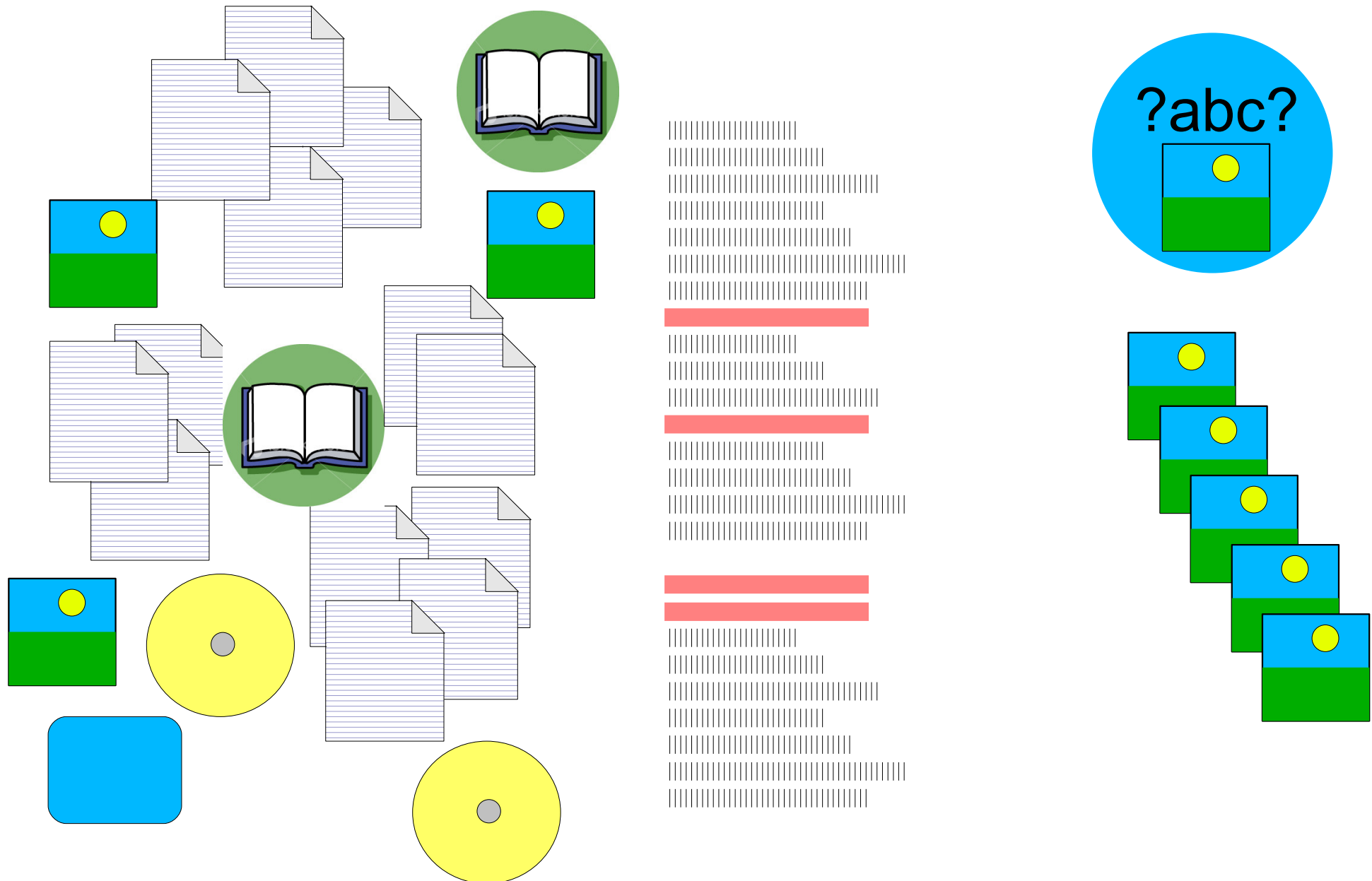
Allgemein: Popularität erhöhen

- Guten Service bieten (schnell, einfach, effektiv, unaufdringlich)
- Nutzende auf der Seite halten und den Zugang vereinfachen (toolbar)
- Weitere Angebote, die Nutzende anziehen und binden

Suchergebnisse verbessern:

- **Suchraum einschränken** nach:
 - Sprache
 - Format (Bilder, Video, Maps, Dateiformate, ...)
 - Ort
 - Zeit
 - Server / Domain
 - Art der Dokumente (News, Scholar, Blogs, Produktsuche, Shopping)
 - und Dienste (Fahrplan, Börsenkurse)

Suchraum nach Format einschränken

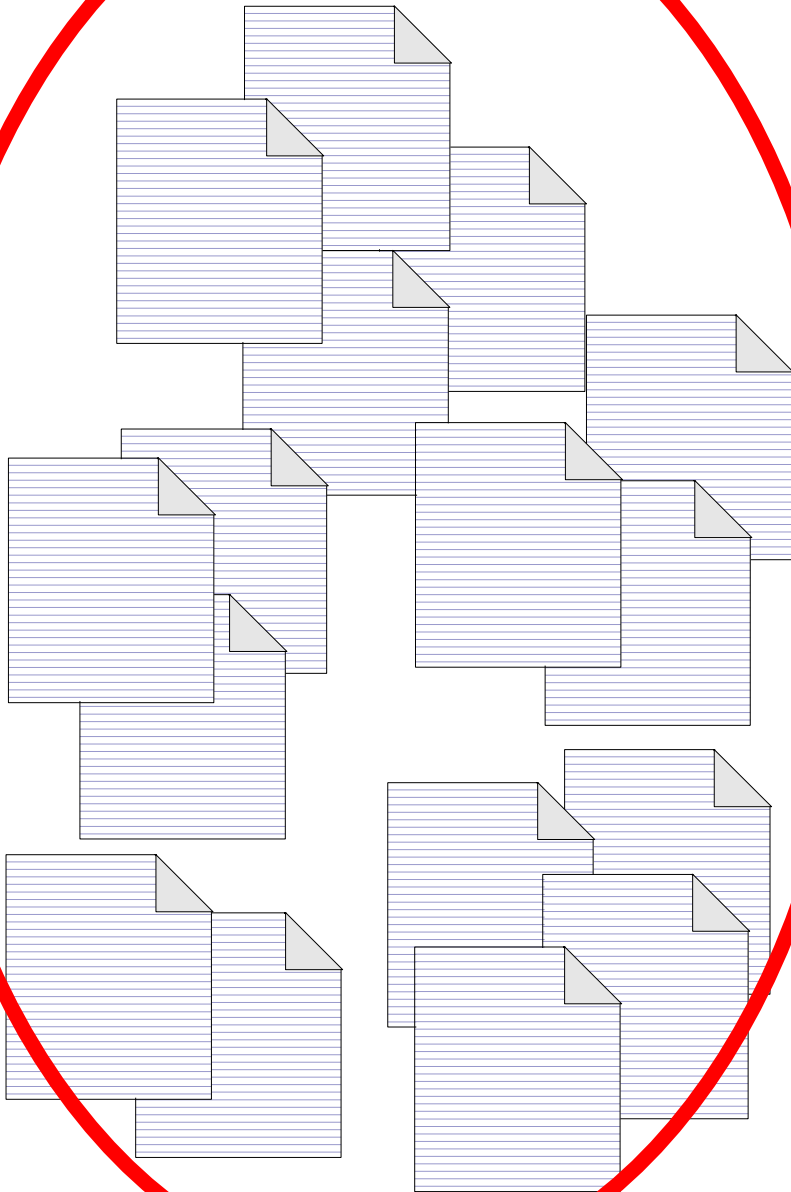


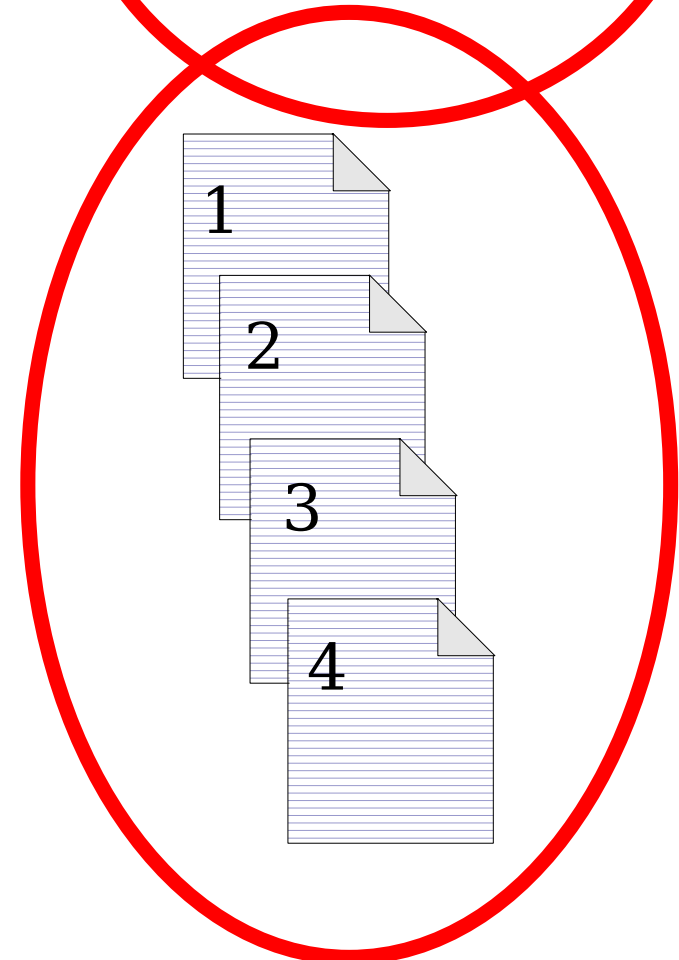
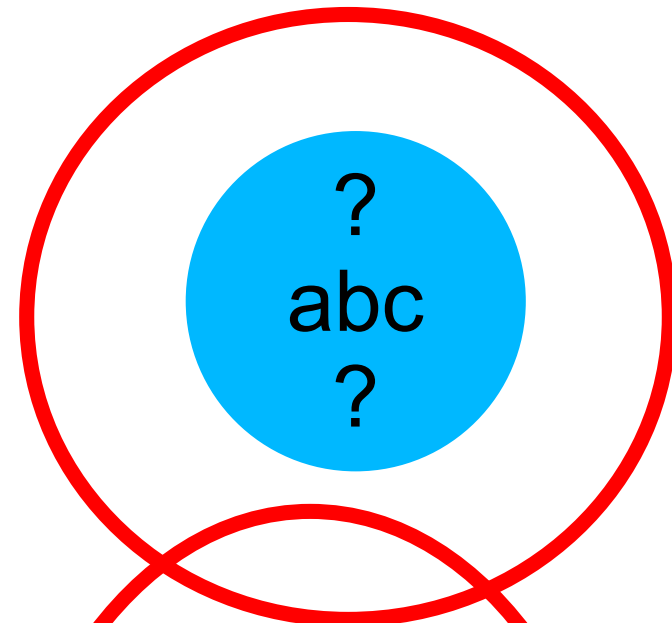
Suchergebnisse verbessern (1)

Verbesserung des **Rankings** durch erweiterte Analyse

- der **einzelnen Seiten** (SEO-Kriterien)
 - Gewichtung von Termen,
 - HTML-Analyse
 - Andere Formate
 - Spamming erkennen
 - der Verzeichnis- und **Serverstruktur**
 - der **Verlinkung**
 - des **Nutzungsverhaltens** auf Seiten
 - Häufigkeit des Besuchs
 - Dauer des Aufenthalts
 - Wege zwischen Seiten
 - Mausbewegungen
- > Google Analytics, (AdSense)

Weitere Möglichkeiten





Suchergebnisse verbessern (2)

- Zusätzliche **Informationen** über die **einzelnen Suchenden** zur Verbesserung der **Anfrage**
 - Suchgeschichte: Login, (Desktop)
 - Interessengebiete: News / Web Alerts, Groups
 - Inhaltsrelevante Dienste: Social Network (orkut), Mail, sites
 - Cloud computing? (text & tabellen)
- Zusätzliche Informationen über **Nutzendengruppen** (Mining)
 - Login, Groups, Social Networks, ...
 - Vergleich von Nutzendenprofilen, Recommendersysteme, ...
- **Anreichern der Ergebnislisten** mit Zusatzinformationen
 - Übersetzungen
 - Karten und Fotos von Orten: Maps, Streetview,
 - Wechselkurse, Fahrpläne, ...

Fazit

Suchmaschinen sind die **Hauptzugangswege** zu den **Web**-Inhalten

Sie **verdecken** die enorme **Heterogenität** des Web

Dazu werden immer **komplexere Methoden** verwendet, die sich auf sehr **große Datenmengen** stützen

Die populärste Web-Suchmaschine ist **Google**

Sie finanziert sich vor allem über **suchwortgesteuerte Anzeigen**

Suchmaschinen müssen (im Prinzip)

- einen Kompromiss zwischen den Interessen
 - der **Werbekunden** und
 - der **Nutzenden** finden
- **bessere** Ergebnisse als die **Konkurrenz** liefern

Zur **Verbesserung** der Ergebnisse können Daten

- aus **unterschiedlichen Bereichen** dienen, die
- mit **unterschiedlichen Anwendungen** gewonnen werden können

Oft ist es wichtig, dass die Daten einer **Person zugeordnet** werden

Ob die Person dabei **anonym** bleibt, ist oft nicht entscheidend