

Symposium

MEHR ALS STICHWORTE: WIE ARBEITEN DIE SUCHMASCHINEN VON MORGEN?

INFORMATIONSTECHNISCHE, POLITISCHE UND
ÖKONOMISCHE PERSPEKTIVEN

Einführung

Friedemann Mattern

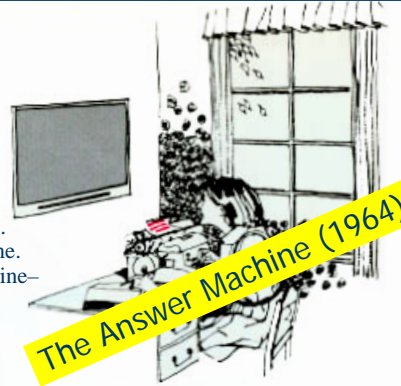
Zur Bedeutung von Suchmaschinen

- > Das **Internet** ist innerhalb weniger Jahre zu einer bedeutenden, nahezu unverzichtbaren **Infrastruktur unserer Gesellschaft** geworden; im geschäftlichen wie auch im privaten Umfeld
- > **Suchmaschinen** sind nach E-Mail die beliebteste und wichtigste Internetanwendung; wer das Internet nutzen will, kommt ohne sie nicht aus; sie bilden die **Universalschnittstelle** zur digitalen Welt
- > Suchmaschinen haben sich zu einem **selbstverständlich genutzten „professionellen“ Rechercheinstrument** entwickelt (Journalisten, Studenten, Wissenschaftler,...)
- > Welche Informationen eine Suchmaschine liefert, liefern soll oder liefern darf, wie das „Ranking“ der Suchergebnisse erfolgt etc. sind mittlerweile Fragen von großer **ökonomischer** und zunehmend auch von **kultureller** und **politischer** Bedeutung

Die Traummaschine von 1964

The Answer Machine

You're doing your homework.
You're stuck and you need some answers.
So you get help from your answer machine.
On a table next to you is part of the machine—
a typewriter keyboard.
When you push the correct keys,
your questions will be answered
on a screen on the wall.



3

Die Traummaschine von 1964

You might ask (using the keyboard):

WHO INVENTED THE PHONOGRAPH?

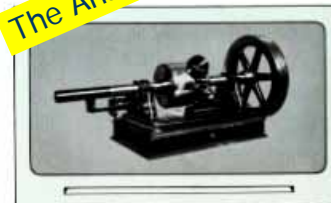
WHAT DID EDISON'S FIRST
PHONOGRAPH LOOK LIKE?

Someday, you might have
an answer machine
that could do all those things

And the machine would answer:

THOMAS A.
EDISON

The Answer Machine (1964)



4

acatech
Wissen schafft Zukunft

Die Antwortmaschine 40 Jahre später...

[Web](#) [Bilder](#) [Groups](#) [News](#) [Mehr »](#)
 Google [Erweiterte Suche](#)
[Einstellungen](#)

Web

[In October of 1877 Edison invented the phonograph - \[Diese Seite übersetzen \]](#)
 In October of 1877 Edison ~~invented the phonograph~~ - several months later in 1878 he made his first production model and this is it! ...
www.rickford.com/edison.htm - 4k - [Im Cache](#) - [Ähnliche Seiten](#)

[Thomas Edison - The Inventions of Thomas Edison - \[Diese Seite übersetzen \]](#)
 He toured the country with the tin foil **phonograph**, and was invited to the ... C. Francis Jenkins and Thomas Armat **invented** a film projector called the ...
inventors.about.com/library/inventors/bledison.htm - 43k - [Im Cache](#) - [Ähnliche Seiten](#)

5

acatech
Wissen schafft Zukunft

Die Antwortmaschine 40 Jahre später...

[Web](#) [Bilder](#) [Groups](#) [News](#) [Mehr »](#)
 Google

Bilder Angezeigt:



Edison's first phonograph
320 x 230 Pixel - 10k - jpg
www.tonalsoft.com



Thomas Edison and George Eastman
450 x 587 Pixel - 48k - jpg
www.edwardsamuels.com

6

Phantastische Leistung von Informatik und Informationstechnik

Suchmaschinen-Historie

- > **1990 (vor dem WWW) „Archie“ und „Veronica“:**
Automatischer Katalog verfügbarer Dokumente auf ftp-Servern
- > **1993 „World Wide Web Wanderer“ / „Webcrawler“:**
Erste Programme „durchlaufen“ das WWW
- > **Ab 1994:** Web-Suchmaschinen wie **AltaVista, Excite, Lycos, Yahoo, HotBot** etc. entstehen



- > **1998:** **Google**

7

- > **Ab ca. 2000: Suchmaschinen ubiquitär**



8



 Wissen schafft Zukunft

> Ab 2004: „googeln“ im Duden



[Web](#)
[Bilder](#)
[Groups](#)
[News](#)
[Mehr »](#)

Suche:  Das Web  Seiten auf Deutsch  Seiten

Web

Meinten Sie: [zweimal täglich googeln](#)

9



 Wissen schafft Zukunft

> Ab 2004: „googeln“ im Duden



[Web](#)
[Bilder](#)
[Groups](#)
[News](#)
[Mehr »](#)

Suche:  Das Web  Seiten auf Deutsch  Seiten  der Schweiz

Web

Meinten Sie: [mutterschaftsurlaub](#)

- > Semantik?
- > Algorithmische Ideologie?

10

Suchanfragen pro Tag

- > 1999: ½ Million
- > 2000: 100 Millionen
- > 2001: 150 Millionen
- > 2003: 200 Millionen
- > 2005: 2,3 Milliarden
- > 2007: 7 Milliarden (ca.)

Häufigste Suchbegriffe

2006 (Google)

1. bebo
2. myspace
3. world cup
4. metacafe
5. radioblog
6. wikipedia
7. video
8. rebelde
9. mininova
10. wiki

2004 (Google)

1. Britney Spears
2. Paris Hilton
3. Christina Aguilera
4. Pamela Anderson
5. chat
6. games
7. Carmen Electra
8. Orlando Bloom
9. Harry Potter
10. mp3

11

Germany – Top Gaining Queries: April 2007 (Google)

- | | | |
|---|--|---|
| 1. <u>torrie wilson</u> | 6. <u>igel</u>
(Medical services not covered by health plans and paid for by the patient) | 11. <u>tim mälzer</u>
(popular TV cook) |
| 2. <u>minigames</u>
(Online Games for free) | 7. <u>rostock</u>
(City in Germany) | 12. <u>schäferhund</u>
(German shepherd) |
| 3. <u>ferienwohnung</u>
(vacation home) | 8. <u>kalender 2008</u>
(calendar 2008) | 13. <u>löwenzahn</u>
(popular telecast for children) |
| 4. <u>del</u>
(short for Deutsche Eishockey Liga = German Hockey League) | 9. <u>facebox.com</u>
(social network for young people) | 14. <u>stammbaum</u>
(family tree) |
| 5. <u>valentino rossi</u>
(Italian motorbike racer) | 10. <u>ikea.de</u> | 15. <u>wallstreet online</u>
(finance portal) |

12

Suchmaschinenmarkt USA, März 2007

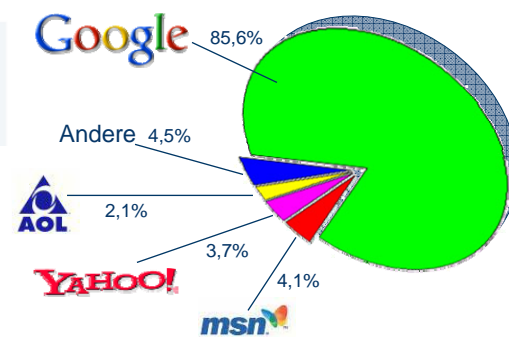
- > Oligopol von drei US-Anbietern:
Google, Yahoo und MSN / Windows Live



- > Web-Abdeckungsgrad: 50% – 80% („surface web“: ca. 15 Milliarden Seiten)
- > Unsichtbares „deep web“ ca. 100 – 500 mal größer?

13

Suchmaschinenmarkt Deutschland, Juli 2006



- > Quasi-Monopol, allerdings ohne Nutzungszwang

14

Suchmaschinen-Macht?

- > **Ranglistenplatz** bei Suchanfragen entscheidend nicht nur für Internetshops, sondern zunehmend auch für „klassische“ Firmen
 - Rangplatz, den eine Seite gewinnt, muss eine andere verlieren
- > **Relevanz-Kriterien** der Index-Erstellung und des Rankings sind Firmengeheimnisse
- > **Risiko des Machtmissbrauchs** durch Marktbeherrschung
 - **Ungleichbehandlung** von Web-Seiten, Institutionen, Anzeigekunden,...?
 - **Recht auf Auflistung** in den Ergebnislisten?
 - **Zensur** / Selbstzensur / Manipulation?
- > **Informationsvorsprung** durch Data-Mining von Nutzerfragen
 - Nutzervorlieben, Trends, Aktienkurse,...

Regulierung?
> Jugendschutz,
> Datenschutz,
> Haftbarkeit,...

15

Wirtschaftliche Bedeutung?

- > **Suchmaschinenbetreiber** als Wirtschaftsfaktor
- > Auf Suchergebnissen **aufbauende Dienstleistungen**
- > Bedeutung von Suchergebnissen für **andere Geschäftsfelder**
- > Alle großen Suchmaschinen sind **US-dominiert**; ebenso andere wichtige New-Economy-Firmen wie **Amazon** oder **eBay** und Web 2.0-Plattformen wie **MySpace.com** oder **Flickr**
 - Haben Deutschland bzw. Europa einen wichtigen Wirtschaftszweig verpasst?
 - Europa nur als Nutzer amerikanisch geprägter Wissensinhalte, Ansichten, Kultur...?
 - Wer digitalisiert „die Welt“?
 - Wem gehören Daten und Wissen?

Digitalisierung
des Weltwissens

16

Medien, Kultur, Gesellschaft?

Google's just seven years old, but it's hard to imagine life before instant search.
-- Joanna Glasner, Wired

- > Wie **beeinflussen** Suchmaschinen das Lernen, das Leben ...?
 - Ultimatives Wissenstool?
 - Ersatz für Bibliotheken?
 - Copy&Paste-Kultur?
 - Beeinflussung von Meinungen?
 - Weltbild erweitert, eingeengt oder schief?
- > Fundamentalkritik von **Jean-Noël Jeanneney** (Präsident der französ. Nationalbibliothek und Autor des Buches „Googles Herausforderung“)
 - „Der Markt steht nicht über Nation und Staat; Staat und Nation müssen den Markt überragen“
 - „Unvermeidliche amerikanische Ichbezogenheit“
 - Plädoyer für eine **europäische** digitale Bibliothek (und eine europäische / französische Suchmaschine)

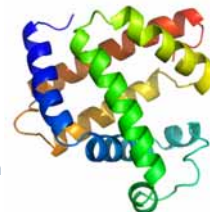


→ Quaero
→ Theseus (D)

17

Informationstechnik für zukünftige Suchmaschinen

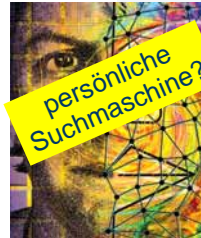
- > „**Neue Daten**“ im Internet: Fotos, Videos, gescannte Bücher, medizinische Datenbanken, sensorgenerierte Realzeitdaten
 - über die Umwelt, den Zustand von Produkten, den momentanen Ort von Dingen, die eigene Gesundheit,...
- > Die Suche in solchen semantisch komplexen Datenbereichen erfordert innovative Mechanismen („**intelligente / semantische Suche**“)
- > Beispiel: Suche in **Bedeutungsklassen** statt Suchen identischer Dinge („letter-by-letter, pixel-by-pixel“)
 - z.B. Forschungsprojekte von Prof. Burkhardt, Freiburg: Ähnlichkeitssuche bei Briefmarken bzw. Proteinfaltungen



18

Ultimative Herausforderungen

- > Aus Daten nützliche **Information, Wissen** und **Wert** generieren
- > **Persönliche Suchmaschine:**
 - Genau auf meine automatisch erkannte aktuelle Situation zugeschnitten
 - Maschinelles Lernen
 - Dienstleister für private Daten und Wissensbedürfnisse



Mit der Speicherung von Wissen in digitalen Dokumenten sind wir zum ersten Mal seit Erfindung der Schrift darauf angewiesen, Maschinen zu besitzen, die in der Lage sind, uns Zugang zum gespeicherten Wissen zu verschaffen.
-- Christine Pläß

19

Programm – Vormittag

11:00 Suchmaschinen – Herausforderung für Medienpolitik, Medienbusiness und Journalismus

Marcel Machill, Universität Leipzig


Von Suchmaschinen zu Antwortmaschinen: Semantische Technologien und Benutzerpartizipation im Web 3.0

Wolfgang Wahlster, DFKI

Technologische Herausforderungen an eine global operierende Websuchmaschine

Thomas Hofmann, Google

12:45 MITTAGSPAUSE




acatech
Wissen schafft Zukunft

MARCEL MACHILL / MARKUS BEILER
(Hrsg.)
DIE MACHT DER SUCHMASCHINEN / THE POWER OF SEARCH ENGINES
2007, ca. 280 S., Broschur,
213 x 142 mm
ISBN 3-938258-33-0
ISBN 978-3-938258-33-0

Herbert von Halem Verlag

- > 21 Beiträge von 29 internationalen Experten
 - > Suchmaschinenregulierung und -ökonomie
 - > Suchmaschinen und Journalismus
 - > Qualität von Suchmaschinen und Nutzerverhalten



acatech
Wissen schafft Zukunft

Programm – Vormittag

11:00 Suchmaschinen – Herausforderung für Medienpolitik, Medienbusiness und Journalismus
Marcel Machill, Universität Leipzig

Von Suchmaschinen zu Antwortmaschinen: Semantische Technologien und Benutzerpartizipation im Web 3.0
Wolfgang Wahlster, DFKI

Technologische Herausforderungen an eine global operierende Websuchmaschine
Thomas Hofmann, Google

12:45 MITTAGSPAUSE

Programm – Nachmittag

**14:00 Wirtschaftliche Aspekte der Suchtechnologie –
es geht um mehr als nur Suche!**

Stefan Wess, empolis

**Auswirkungen von Suchmaschinen auf den
Betrieb von Online-Angeboten**

Michael Buschmann, Deutsche Telekom

Volltextsuchdienste – was darf/soll wie hinein?

Rainer Kuhlen, Universität Konstanz

15:05 KAFFEPAUSE

15:20 Auf dem Weg in die Google-Gesellschaft?

Rudi Schmiede, TU Darmstadt

15:35 Podiumsdiskussion (bis ca. 16:45)

Vielen Dank!

Prof. Dr. Friedemann Mattern
Institut für Pervasive Computing
ETH Zürich
CH-8092 Zürich
Schweiz

Tel.: +41 44 632 05 36
Fax.: +41 44 632 16 59
E-mail: mattern@inf.ethz.ch