



Zürcher  
Hochschule  
Winterthur

InIT

# Evaluation der Suchfunktion von Schweizer Unternehmens-Websites: Einleitung

Dr. Martin Braschler

Dozent, Institut für angewandte Informationstechnologie  
Zürcher Hochschule Winterthur

30.11.2006

# Inhalt

- Einführung
- Ziele
- Einordnung
- Abgrenzung
- Retrievalproblem
- Evaluationsraster
- Testdurchführung
- Analyse
- Boxplots

# Informationsflut

- Grosse Unternehmens-Websites bieten heute eine so umfangreiches Informationsangebot, dass leistungsfähige Zugriffspfade essentiell sind
- Ausgefeilte Navigation ist wichtig
- Der Nutzer erwartet aber insbesondere auch eine leistungsfähige Suchfunktion
- Eine Suche erlaubt es, "ad-hoc" ein bestimmtes Informationsbedürfnis zu adressieren

# Unternehmens-Websites als Kommunikationsinstrumente

- Unternehmens-Websites dienen heute als Kommunikationsinstrumente der Informationsanbieter
  
- Anlaufstelle für Informationen über
  - Neuheiten
  - Dienstleistungen
  - Produkte
  - Kontaktdetails, u.a.
  
- Konzentration der Studie auf Unternehmen und öffentliche Behörden

# Kommunikation mit der Website

- Effektiver und effizienter Zugriff auf gesuchte Information ist zentral
  - relevante, und nur relevante Information
  - schnell, unkompliziert, transparent
- Der Nutzer initiiert die Kommunikation mit der Websites im Wesentlichen selbst, er kann diese, wenn seine Anforderungen unerfüllt bleiben, jederzeit abbrechen

# Ziele

- Den State-of-the-Art in Sachen Suchfunktion in Unternehmens-Websites ermitteln
- Klassierung in gute, mittlere, schlechte Site-Suchen
- Bedürfnisse im Mittelpunkt, nicht die Implementation (technische Funktionen)
- Identifikation der besten Ansatzpunkte zur Verbesserung der Suchfunktion
- Erarbeiten eines vollständigen Evaluationsraster, wissenschaftlich fundiert

# Internet-Suchdienste

- Das Informationsangebot weltweit wächst stetig
- (Lyman & Varian, 2003) schätzen einen jährlichen Zuwachs von 5 Exabytes an Daten im Jahr 2002, die Gesamtmenge der Daten hat sich zwischen 1999 und 2002 verdoppelt
- Internet-Suchdienste wie Google, Yahoo, MSN sind heute populär
- Ermöglichen Zugriff auf ein BREITES Angebot mit Stichworten, unabhängig der Verteilung und der Charakteristiken der erschlossenen Information
- Müssen zwischen mehreren Anbietern der gleichen/widersprüchlichen Information gewichten

# Suchfunktion für Unternehmens-Websites

- Die Situation für Unternehmens-Websites stellt sich grundlegend anders dar.
- Die Nutzer sind möglichst ohne Umwege auf sehr spezifische Information zu führen
- Nutzer haben spezifische Informationsbedürfnisse, sind auch ausbeuteorientiert
- Die angebotene Information kann geeignet vorstrukturiert werden.



# Vertikale vs. Horizontale Suche

|                                 | <b>Suchfunktion für Unternehmens-Websites (vertikale Suche)</b> | <b>Internet-Suchdienste (horizontale Suche)</b>          |
|---------------------------------|---|--|
| <b>Gewichtungsmechanismen</b>   | Auf Präzision und Ausbeute optimiert                            | Präzisionsorientiert                                     |
| <b>Erschlossene Information</b> | Homogene Qualität, oft hierarchisch strukturiert                | Breit, heterogene Qualität, unterschiedlich strukturiert |
| <b>Sucheinschränkungen</b>      | Orientieren sich am Informationsangebot, Unternehmensstruktur   | Wenig vorhanden  |
| <b>Suchassistenten</b>          | Thesauri, Kategoriebäume  | Typischerweise keine                                     |
| <b>Informationsangebot</b>      | Tausende von Seiten   | Milliarden von Seiten                                    |

# Suchfunktion als Retrievalapplikation

- Die Studie stellt die Suchfunktion eines Unternehmens-Websites als Kombination von Software, Konfiguration und Informationsangebot ins Zentrum
- Die Suchfunktion stellt damit eine "Retrievalapplikation" dar, die eine kommunikative Aufgabe für den Betreiber erfüllt

# Abgrenzung

- Kein Produktvergleich
- Kein reiner Systemvergleich
- Keine Listen von technischen Funktionen
- Kein Test von "breiten", "horizontalen" Internetsuchmaschinen
- Die ermittelten Leistungswerte sind nicht auf "artfremde" Szenarien im gleichen Unternehmen übertragbar (Intranet)

# Abgrenzung

- Kein Ranglisten von Unternehmen
- Keine Awards
- Keine Interpretation von "absoluten" Einzel-Leistungswerten (85 von 100 Punkten auf der Evaluationsskala)
- Es gibt fundamentale, wissenschaftliche Gründe, die ein solches Vorgehen verbieten
- Die Resultate sind anonymisiert

# Das Retrievalproblem

- Nutzer suchen auf "fremden" Daten, kennen im Allgemeinen die Datenbasis nicht
  - Informationsbedürfnisse sind vage, auch durch den Nutzer selbst schlecht oder unvollständig verstanden
  - Interpretation der gefundenen Information variiert von Nutzer zu Nutzer: unterschiedliches Hintergrundwissen, Präferenzen, Anwendungszweck
  - Die gleiche Information kann auf vielfältige Art dargestellt werden
- Kein Datenbankszenario

# Das Retrievalproblem

- "Das Auffinden von möglichst viel relevanter Information bei gleichzeitigem Minimieren der ebenfalls gelieferten irrelevanten Information"
- Relevanz ist subjektiv
- Es gibt kein "perfektes" Retrievalergebnis

# Konsequenzen für die Evaluation

- Absolute Bewertung von Suchresultaten im wissenschaftlichen, objektiven Sinn ist sehr schwierig
- Effektivität von Retrievalsystemen wird z.B. nach dem Cranfield-Paradigma bestimmt: Systeme werden verglichen, relative Aussagen gemacht
- Je mehr Systeme in die Evaluation einbezogen werden, desto aussagekräftiger die Evaluation (in dieser Studie: 56 Unternehmens-Websites)

# Konsequenzen für die Evaluation

- Viele Einzelkriterien erhöhen die Stabilität der Aussagen
- Relative Aussagen (Rangfolge der Probanden) sind stabil
- Zuverlässige Aussagen über die typische (mittlere) Leistung sind möglich



# Vorbilder

- Die grossen Evaluationskampagnen für Retrievalsysteme und –methoden gehen nach diesem Muster vor
- TREC ([trec.nist.gov](http://trec.nist.gov)), seit 1992
- CLEF ([www.clef-campaign.org](http://www.clef-campaign.org)), seit 2000
- NTCIR ([research.nii.ac.jp/ntcir](http://research.nii.ac.jp/ntcir)), seit 1999

# Evaluationsraster

- Die Evaluation wurde strukturiert in vier Hauptkriterien
  1. Suchindex
  2. Anfrage- / Dokumentenvergleich
  3. Benutzerinteraktion
  4. Suchergebnisse
  
- Die Hauptkriterien wurden in Subkriterien unterteilt, die wiederum aus Einzeltests bestehen
  
- Die einzelnen Resultate wurden aggregiert. Gewichtung gemäss Evaluationsaufbau

# Einzeltests

- Die Einzeltests wurden überwiegend manuell durchgeführt
- Die Tester befolgten für jeden Einzeltest ein detailliertes Testprotokoll
- Die Codierung der Resultate wurde ebenfalls vorgegeben
- Ausgewählte Tests wurden automatisiert durchgeführt (Vollständigkeit, Performance)

# Pre-Test

- Das Testprozedere wurde in einem Pre-Test geprüft
- Pre-Test von HTW Chur, Namics & Eurospider durchgeführt
- Anpassung der Kriterien: einige Einzeltests haben sich als undurchführbar erwiesen, andere Einzeltests wurden neu hinzugefügt
- Die Reproduzierbarkeit wurde überprüft

# Getestete Websites



- Die Evaluation umfasst 56 Unternehmens-Websites
- Ein Unternehmen bat, nicht genannt zu werden

www.abb.ch

■ www.credit-suisse.com

■ www.national.ch

■ www.synthes.com

www.admin.ch

■ www.css.ch

■ www.novartis.ch

■ www.tkb.ch

www.allianz-suisse.ch

■ www.ethzh.ch

■ www.pax.ch

■ www.ubs.com

www.bankcoop.ch

■ www.generali.ch

■ www.phonak.ch

■ www.unibe.ch

www.baselland.ch

■ www.gkb.ch

■ www.raiffeisen.ch

■ www.unique.ch

www.bcv.ch

■ www.helvetiapatria.ch

■ www.railaway.ch

■ www.valiantbank.com

www.be.ch

■ www.holcim.ch

■ www.ringier.ch

■ www.vbs.admin.ch

www.bekb.ch

■ www.juliusbaer.ch

■ www.roche.com

■ www.visana.ch

www.bfs.admin.ch

■ www.kaba.ch

■ www.sgkb.ch

■ www.zh.ch

www.bkb.ch

■ www.lonza.ch

■ www.sunrise.ch

■ www.zhwin.ch

www.blkb.ch

■ www.lukb.ch

■ www.suva.ch

■ www.zkb.ch

www.bs.ch

■ www.migrosbank.ch

■ www.swiss.com

■ www.zug.ch

www.cablecom.ch

■ www.mobi.ch

■ www.swisslife.ch

■ www.zugerkb.ch

www.converium.com

■ www.nab.ch

■ www.swissre.com

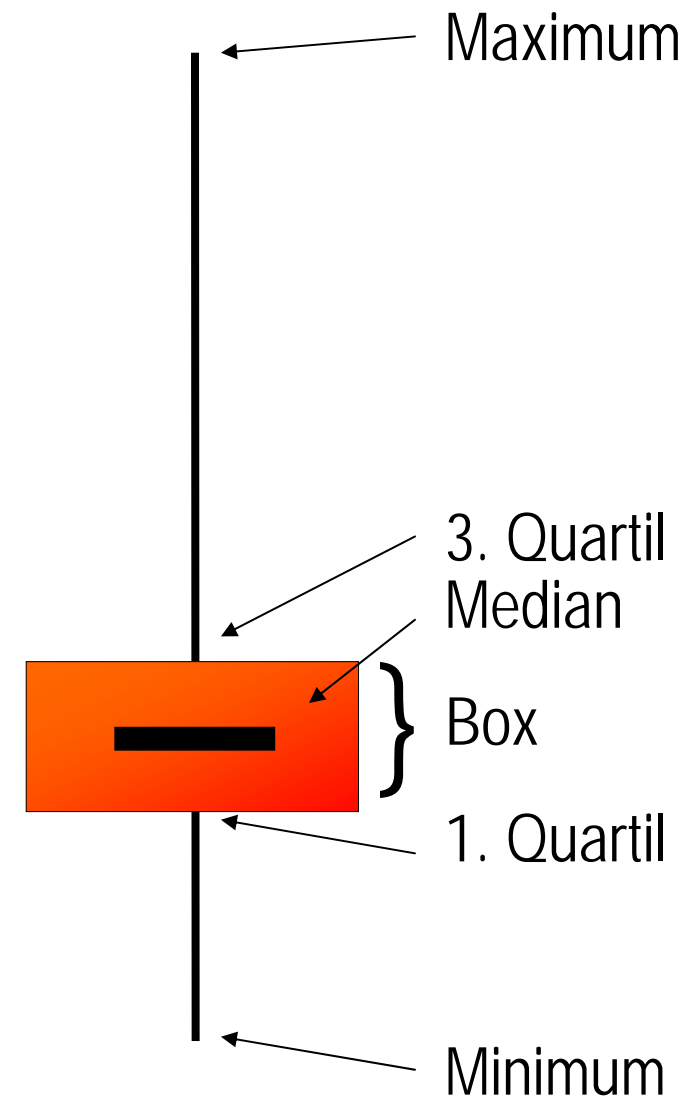
# Durchführung

Tester:

- 2 Studenten Zürcher Hochschule Winterthur
- Mitarbeiterin Namics AG
- Automatisierte Tests: Eurospider
  
- Tests von „aussen“: keine Anpassungen der Websites auf Tests möglich
  
- Aufwand: 4-6 Std. pro Website

# Grafische Darstellung der Resultate

- Die einzelnen Leistungswerte werden nach der Normalisierung mit Hilfe von Boxplots visualisiert
- 1. Quartil = die Resultate 75% aller Systeme sind besser als dieser Wert
- Median = die mittlere Leistung
- 3. Quartil = die Resultate 25% aller Systeme sind besser als dieser Wert
- Die Box repräsentiert die Leistung von exakt der Hälfte der Systeme



# Grafische Darstellung der Resultate

- Die Box stellt die Streuung der Resultatwerte dar (für die "typischen" Probanden)
- Die Lage des Median in der Box ist ein Hinweis auf die Schiefe der Verteilung
- Werte weiter als  $1.5 \cdot$  Interquartilsabstand (=Höhe der Box) von der Box gelten als "schwache Ausreisser"
- Werte  $> 3 \cdot$  Interquartilsabstand von der Box gelten als "extreme Ausreisser"

