



Zürcher  
Hochschule  
Winterthur

InIT

# Praxisstudie: Evaluation der Suchfunktion deutscher Unternehmens-Websites Einleitung

Dr. Martin Braschler

Dozent, Institut für angewandte Informationstechnologie  
Zürcher Hochschule Winterthur

5.7.2007

# Inhalt

- Hintergrund
- Ziele
- Einordnung
- Abgrenzung
- Retrievalproblem
- Evaluationsraster
- Testdurchführung
- Analyse
- Darstellung (Boxplots)

## Hintergrund: Informationsflut

- Das Informationsangebot weltweit wächst stetig
- (Lyman & Varian, 2003) schätzen einen jährlichen Zuwachs von 5 Exabytes an Daten jeglicher Herkunft im Jahr 2002
- Die Gesamtmenge der Daten hat sich zwischen 1999 und 2002 verdoppelt

# Hintergrund: Informationsflut

- Grosse Unternehmens-Websites bieten heute ein 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

# Hintergrund: Unternehmens-Websites als Kommunikationsinstrumente

- Unternehmens-Websites dienen heute als Kommunikationsinstrumente/Anlaufstelle für Informationen der Informationsanbieter
- Daher: Konzentration der Studie auf Unternehmen und öffentliche Behörden
  
- Effektiver und effizienter Zugriff auf gesuchte Information ist zentral
  - relevante Information
  - schnell, unkompliziert, transparent
- Die Studie misst diese Aspekte durch verschiedene (Haupt-)kriterien

# 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 Evaluationsrasters, wissenschaftlich fundiert

# Internet-Suchdienste

- 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
- Kämpfen mit vorsätzlicher "Sabotage" durch Spammer

# 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>Suchassistentz</b>	Thesauri, Kategoriebäume	Typischerweise keine
<b>Informationsangebot</b>	Tausende von Seiten	Milliarden von Seiten

# Suchfunktion als Retrievalapplikation

- Die Studie stellt die Suchfunktion einer 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 Produktevergleich
- 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

- Keine Ranglisten von Unternehmen
- Keine Awards
  
- Keine Interpretation von "absoluten" Einzel-Leistungswerten (85 von 100 Punkten auf der Evaluationskala)
  
- 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: 35 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: TREC, CLEF, NTCIR



# 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 werden in der Nachanalyse automatisiert durchgeführt werden (Vollständigkeit, Performance)

# Einzeltests

- Die Resultate der Tests werden mit geeigneter Abstufung festgehalten
  
- Führt ein Test zu keinem befriedigenden Ergebnis, so wird unterschieden zwischen
  - Funktion nicht vorhanden (berechtigterweise/fälschlicherweise)
  - Test nicht auswertbar (Voraussetzungen fehlen)
  - Funktion fehlerhaft (z. B. Abstürze etc.)
  
- Nicht jeder Test ist für jede Unternehmens-Website zutreffend

# Einzeltests

- Die Resultate der Einzeltests werden numerisch codiert
- Die Leistungswerte eines Systems für ein Subkriterium basieren auf einem gewichteten Mittel der normalisierten Werte der Einzeltests
- Die Leistungswerte für ein Hauptkriterium basieren auf einem gewichteten Mittel der normalisierten Werte der Subkriterien

## Studie Schweizer Unternehmens-Websites

- Das Testprozedere basiert auf einer umfangreichen Studie zu den Merkmalen Schweizer Unternehmens-Websites von 2006
- Diese Studie wurde von HTW Chur, namics, Eurospider und der Zürcher Hochschule Winterthur durchgeführt
- Anpassung der Kriterien: einige Einzeltests haben sich als undurchführbar erwiesen, andere Einzeltests wurden neu hinzugefügt. Vor allem die Behandlung von Grenzfällen wurde stark verfeinert
- Die Reproduzierbarkeit wurde überprüft

# Google-Baseline-Test

- Was passiert, wenn man Google statt einer angepassten Suchlösung einsetzt? ("das kann Google ja eh besser")
- Auswertung der Suche mit Google und Site-Einschränkung (www.google.de site:.....)
- Drei mal vollständig ausgewertet
- 13mal nur Site-spezifische Tests ausgewertet
- Aus diesen Resultaten werden eine "mittlere" Leistung und Streuung ermittelt

# Berücksichtigte Branchen

- Unternehmen der folgenden Branchen wurden eingeladen, ihre Websites analysieren zu lassen:
- Banken
- Versicherungen
- Chemie
- Pharma
- sonstige verarbeitende Industrie
- Telekommunikation
  
- und weitere interessierte Unternehmen

# Getestete Websites

- Die Evaluation umfasst 35 Unternehmens-Websites
  - Ein Unternehmen bat, nicht genannt zu werden
- |  |  |
|--|--|
| ■ <a href="http://www.afriso.de">www.afriso.de</a>                           | ■ <a href="http://kkh.de">kkh.de</a>   |
| ■ <a href="http://www.amv.de">www.amv.de</a>                                 | ■ <a href="http://www.lbs.de/bw">www.lbs.de/bw</a>                                     |
| ■ <a href="http://www.awd.de">www.awd.de</a>                                 | ■ <a href="http://www.lbs.de/nord">www.lbs.de/nord</a>                                 |
| ■ <a href="http://suchergebnisse.barmenia.de">suchergebnisse.barmenia.de</a> | ■ <a href="http://www.mahle.com">www.mahle.com</a>                                     |
| ■ <a href="http://www.bbraun.de">www.bbraun.de</a>                           | ■ <a href="http://www.man.de">www.man.de</a>   |
| ■ <a href="http://www.beiersdorf.de">www.beiersdorf.de</a>                   | ■ <a href="http://www.mdr.de">www.mdr.de</a>   |
| ■ <a href="http://www.deka.de">www.deka.de</a>                               | ■ <a href="http://www.mw-zander.com">www.mw-zander.com</a>                             |
| ■ <a href="http://www.dortmund-airport.de">www.dortmund-airport.de</a>       | ■ <a href="http://www.phoenixcontact.de">www.phoenixcontact.de</a>                     |
| ■ <a href="http://www.ede.de">www.ede.de</a>                                 | ■ <a href="http://www.rofin.com">www.rofin.com</a>                                     |
| ■ <a href="http://www.eppendorf.de">www.eppendorf.de</a>                     | ■ <a href="http://www.sick.de">www.sick.de</a>   |
| ■ <a href="http://www.fraunhofer.de">www.fraunhofer.de</a>                   | ■ <a href="http://www.tk-online.de">www.tk-online.de</a>                               |
| ■ <a href="http://www.glaxosmithkline.de">www.glaxosmithkline.de</a>         | ■ <a href="http://www.t-mobile.de">www.t-mobile.de</a>                                 |
| ■ <a href="http://www.hama.de">www.hama.de</a>                               | ■ <a href="http://de.computers.toshiba-europe.com">de.computers.toshiba-europe.com</a> |
| ■ <a href="http://www.helvetia.de">www.helvetia.de</a>                       | ■ <a href="http://www.victoria.de">www.victoria.de</a>                                 |
| ■ <a href="http://www.hsh-nordbank.de">www.hsh-nordbank.de</a>               | ■ <a href="http://www.volkfuersorge.de">www.volkfuersorge.de</a>                       |
| ■ <a href="http://www.jab-becker.de">www.jab-becker.de</a>                   | ■ <a href="http://www.vpv.de">www.vpv.de</a>   |
| ■ <a href="http://www.kfw.de">www.kfw.de</a>                                 | ■ <a href="http://www.wdr.de">www.wdr.de</a>   |



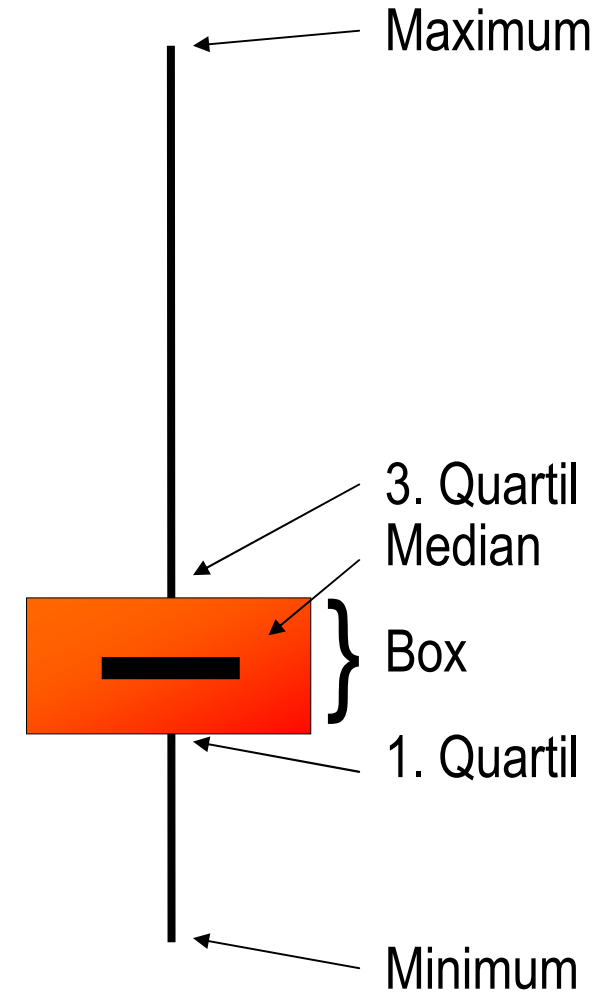
# Durchführung

Tester:

- 5 Tester, Eurospider, HTW Chur, namics, ZHW
- Tests von „aussen“: keine Anpassungen der Websites auf Tests möglich

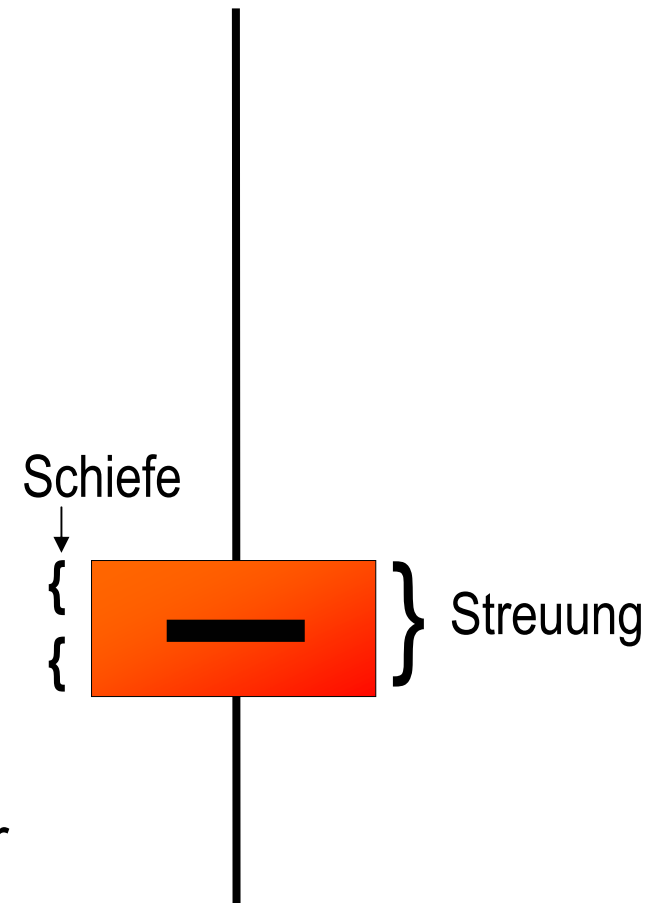
# Grafische Darstellung der Resultate

- 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"



# Darstellung der Google-Baseline

- Ebenfalls als Box-Plots
- Mittels Box-Plots wird die Bandbreite der gemessenen Leistung von Google auf den 16 Websites dargestellt
- Nicht Site-spezifische Tests werden durch den Mittelwert der 3 vollständigen Tests repräsentiert