

HTW Chur
Hochschule für Technik und Wirtschaft

Fachhochschule Ostschweiz
University of Applied Sciences

Churer Schriften zur Informationswissenschaft

Herausgegeben von Josef Herget und Sonja Hierl

Arbeitsbereich
Informationswissenschaft

Schrift 12

Evaluation der Suchfunktion von Schweizer Unternehmens-Websites

Martin Braschler, Josef Herget,
Joachim Pfister, Peter Schäuble,
Markus Steinbach, Jürg Stuker

Chur 2006

Churer Schriften zur Informationswissenschaft

Herausgegeben von Josef Herget und Sonja Hierl

Schrift 12

Evaluation der Suchfunktion

von Schweizer Unternehmens-Websites

Martin Braschler¹, Josef Herget², Joachim Pfister²,
Peter Schäuble³, Markus Steinbach⁴, Jürg Stuker⁴

Verlag: Arbeitsbereich Informationswissenschaft

ISSN: 1660-945X

Chur, November 2006

- 1) Zürcher Hochschule Winterthur
Institut für angewandte Informationstechnologie
Technikumsstrasse 9
8400 Winterthur
brn@zhwin.ch
- 2) Hochschule für Technik und Wirtschaft HTW Chur
Swiss Institute for Information Research (SII)
Ringstrasse/Pulvermühlestrasse 57
7000 Chur
Josef.Herget@fh-htwchur.ch | Joachim.Pfister@fh-htwchur.ch
- 3) Eurospider Information Technology AG
Schaffhauserstrasse 18
8006 Zürich
schauble@eurospider.com
- 4) namics AG
Teufenerstrasse 19
9000 St. Gallen
Markus.Steinbach@namics.com | Juerg.Stuker@namics.com

Abstract

Unternehmen und Behörden in der Schweiz verwenden ihre Websites vermehrt zur Kommunikation mit interessierten Nutzern. Diese Entwicklung fällt zusammen mit einem Umfeld, in dem die Nutzer mit einer steigenden Informationsflut konfrontiert sind. Es ist daher nicht erstaunlich, dass Suchdienste zu den beliebtesten Anwendungen überhaupt auf dem Internet gehören.

Die Ausgangslage für Suchfunktionen auf Unternehmens-Websites („vertikale“ Site-Suchen) stellt sich dabei grundlegend anders dar als bei „horizontalen“ Internet-Suchdiensten. Nicht die möglichst breite Erschliessung von Information vieler verschiedener Websites, sondern der erfolgreiche Dialog mit den Nutzern steht im Vordergrund. Die vorliegende Studie versucht eine umfangreiche Evaluation der Suchfunktionen von Schweizer Unternehmens-Websites. Nicht eine Liste von technischen Funktionalitäten, sondern die Bedürfnisse der Nutzer in ihrer Kommunikation mit der Website standen im Vordergrund.

Es wurden insgesamt 74 Einzeltests grossenteils manuell durchgeführt, die in vier Hauptkriterien gruppiert wurden: Suchindex, Vergleich von Anfragen und Dokumenten, Benutzerinteraktion und Suchergebnisse. Ein wissenschaftlich fundiertes Raster wurde erstellt, in der Tradition der Evaluationen des zugehörigen akademischen Gebietes des „Information Retrieval“. Die Site-Suchen von 54 Schweizer Unternehmen und Behörden wurden der Evaluation unterzogen.

Die Resultate zeigen grosses Potential für Verbesserungen. Dabei wurden kein Ranking der Websites und keine Verleihung von Awards vorgesehen. Vielmehr wurde eine Einteilung in gute, mittlere und schlechte Websites für die einzelnen Kriterien ermöglicht. Die Analyse zeigt denn auch, dass keine Website für alle Kriterien sich jeweils unter den Besten klassieren konnte.

Viele Erkenntnisse betreffen direkt die Kommunikationsfunktion der Website: so konnten viele Suchfunktionen den Anforderungen hinsichtlich Vollständigkeit und Aktualität der Datenbasis, den notwendigen Voraussetzungen für gute Suchergebnisse, nicht genügen. Informationen werden oft nur mit inkorrekten Titeln oder Datumsinformationen zugänglich gemacht. Und nicht zuletzt schnitten viele Websites bei der klassischen Aufgabe des „Information Retrieval“, wo mittels „Informationsanfragen“ durch Nutzer gezielt Information zusammengestellt werden, ungenügend ab.

Die Studie liefert mittels Beispielen und Empfehlungen den richtigen Einstieg, der Suchfunktion von Unternehmens-Website die Aufmerksamkeit zu schenken, die ihrer heutigen Bedeutung als Kommunikationsinstrument gerecht wird.

Inhaltsverzeichnis

| | |
|---|----|
| Vorwort..... | 4 |
| 1. Einleitung..... | 6 |
| 1.1. Websites als Kommunikationsinstrument..... | 6 |
| 1.2. Die Studie | 7 |
| 1.3. Abgrenzung | 8 |
| 1.4. Das Retrievalproblem..... | 9 |
| 1.5. Relevanz..... | 10 |
| 2. Testanordnung und Durchführung..... | 12 |
| 2.1. Evaluations-Raster | 12 |
| 2.2. Untersuchte Websites: | 13 |
| 3. Gesamtresultat | 14 |
| 3.1. Boxplots..... | 14 |
| 3.2. Diskussion der Hauptkriterien..... | 15 |
| 3.3. Gesamtheitliche Betrachtung | 17 |
| 3.4. Allgemeiner Eindruck..... | 17 |
| 4. Suchindex..... | 19 |
| 5. Anfragen und Dokumente vergleichen | 22 |
| 6. Benutzerinteraktion..... | 26 |
| 7. Suchergebnisse..... | 29 |
| 8. Empfehlungen | 32 |
| 9. Kritische Würdigung der Studie | 34 |
| 10. Literaturverzeichnis | 36 |
| 11. Glossar | 38 |

Vorwort

Suchdienste stellen heute im Internet das zentrale Werkzeug zur Informationsbeschaffung dar. Diese horizontalen Internet-Suchdienste wie Google und Yahoo eignen sich für die generelle Suche innerhalb des gesamten Internets.

Auch die Kommunikation zwischen Kunden und Organisationen aller Art erfolgt zunehmend über das Internet. Websites von Unternehmen entwickeln sich zu Kommunikationsdreh-scheiben und sie sind häufig die erste Anlaufstelle, um Informationen über Unternehmen, Angebote, Produkte oder Ansprechpartner zu suchen – und hoffentlich zu finden. Orientierungshilfe liefert vorrangig die Navigation über die Verlinkungen innerhalb der Websites, zumeist wird zusätzlich auch eine Sitemap und eine integrierte Suchfunktion angeboten. Doch wie gut ist diese vertikale Suche geeignet, dem Kunden auch die gewünschte Informa-tion zu liefern?

Aus der Benutzerforschung ist schon seit langem das Phänomen bekannt, dass der Informa-tionssuchende häufig nicht einmal in der Lage ist, sein Informationsproblem exakt zu beschreiben bzw. sein Anliegen genau zu formulieren. Dieses Phänomen wird als „anoma-lous state of knowledge“ bezeichnet. Problemverstärkend stellt sich zudem folgende Frage: Wissen wir denn überhaupt, was auf eine Suchanfrage hin dem Informationssuchenden als relevantes Ergebnis präsentiert wird? Oder wird dies einem zufälligen – uns zumeist auch unbekanntem – Algorithmus der Suchmaschine überlassen? Der Erfolg der Kommunikation, also ob der Kunde die richtige Information tatsächlich erhält, entzieht sich weitestgehend der Kontrolle des Unternehmens!

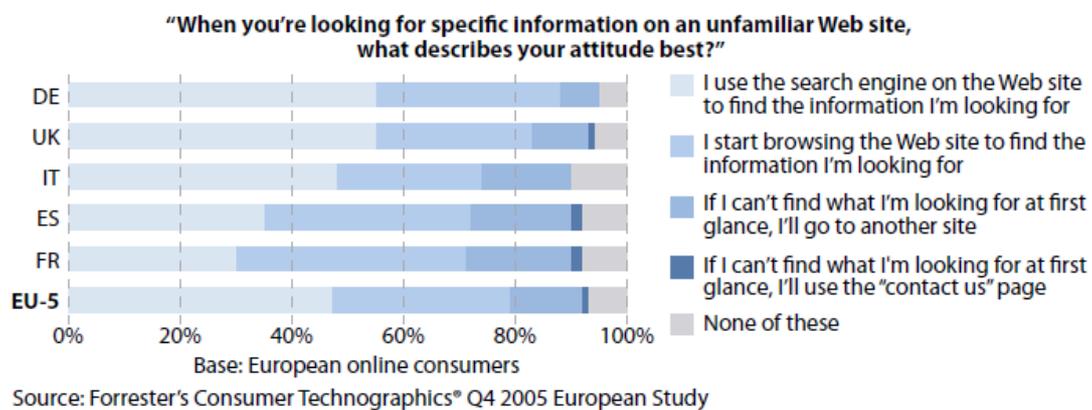
Die Betreiber von immer umfassender und komplexer werdenden Websites müssen sich schlussendlich fragen lassen, ob Sie zum einen überhaupt wissen, was die Interessenten suchen und zum anderen, ob das, was die Suchfunktion als Resultat ermittelt auch das ist, was wirklich kommuniziert werden soll. Die Qualität der Beantwortung der Suchanfrage, dar-um geht es schliesslich, bleibt also häufig dem Zufall überlassen – eine unbefriedigende Situation sowohl für Kunden als auch für Anbieter. Anders als bei persönlichen oder schrift-lichen Kommunikationssituationen, bei denen wir selbstverständlich grossen Wert auf eine konsistente und angemessene Beantwortung von Anfragen legen, scheint dies für die Kom-munikation im Internet nicht zu gelten. Dort ist letztendlich der Suchende für den Erfolg seiner Suchfrage selbst verantwortlich.

Die Studie „Evaluation der Suchfunktion von Schweizer Unternehmens-Websites“ versucht nun folgende Fragen zu klären: Wie gut gelingt es den Nutzern mit der Site-Suche vorhan-

dene Informationen zu ermitteln? Wo gibt es Problembereiche, die eine erfolgreiche Kommunikation stören bzw. verhindern?

Die Bedeutung der Suchfunktion für eine Website wird beispielsweise in Untersuchungen von Forrester Research herausgestellt (vgl. Cremers, 2006). Besonders wenn nach spezifischen Informationen auf unbekanntem Websites gesucht wird, verlassen sich die Nutzer mehr auf die Suchfunktion der Website, als dass sie sich selbst auf den Sites die gesuchten Informationen navigierend erarbeiten (vgl. Abbildung 1). Zudem zeigen die Ergebnisse der Forrester-Umfrage (Cremers, 2006), dass durchschnittlich 13% der Nutzer zu einer anderen Site wechseln, wenn sie ihr Informationsbedürfnis nicht sofort gestillt sehen.

Figure 2 Consumers Turn To Search At Unfamiliar Sites



Source: Forrester Research, Inc.

Abbildung 1: Welche Zugriffsmöglichkeit bevorzugen Nutzer auf unbekanntem Websites?

Die Suchfunktion auf einer Website ist aus Nutzersicht folglich der zentrale Anknüpfungspunkt zur Ermittlung von Informationen aus einem Web-Angebot. Die Anbieter von Informationsangeboten im Web müssen dieser Entwicklung Rechnung tragen, indem sie leistungsfähige und den Erwartungen der Nutzer entsprechende Such-Tools einsetzen.

Zur Diagnose der Leistungsfähigkeit der Suchfunktion wurde in enger Zusammenarbeit der Projektpartner Eurospider, Hochschule für Technik und Wirtschaft (HTW) Chur, namics und der Zürcher Hochschule Winterthur ein Evaluations-Raster erarbeitet. Die vorliegende Studie fasst die Ergebnisse der ersten Evaluationsrunde zusammen und gibt einen generalisierten Überblick, wo die Stärken und Schwächen der getesteten Site-Suchen liegen. Sie soll dazu anregen, den Stellenwert der Suchfunktion auf einer Website neu zu überdenken und die Suche als ein Mittel zur effektiven Kommunikation zwischen Anbieter und Nutzer wahrzunehmen.

1. Einleitung

Das Informationsangebot grosser Unternehmens-Websites ist heute so umfangreich, dass Nutzer auf leistungsfähige Zugriffspfade angewiesen sind. Neben einer ausgefeilten Navigation

erwartet der Nutzer heute im Allgemeinen eine leistungsfähige Suchfunktion¹. Diese Site-Suche ermöglicht es, „ad-hoc“ ein bestimmtes Informationsbedürfnis auszuwerten. Dabei soll die Suche mittels Stichwörtern möglich sein, d.h. weitgehend unabhängig zur vorgängigen Strukturierung der Information auf der Website.

Unternehmens-Websites dienen dabei als Kommunikationsinstrumente der Informationsanbieter. Diese dienen als Anlaufstelle für interessierte Personen, die sich beispielsweise über Neuheiten, Dienstleistungen, Produkte, Adressdetails und anderes der Anbieter informieren können. Anbieter von Websites in der Schweiz sind vielfältig. Die Studie konzentriert sich dabei ausschliesslich auf professionelle Anbieter. Dabei handelt es sich um Firmen und öffentliche Behörden.

1.1. *Websites als Kommunikationsinstrument*

Sieht man Websites nun als Plattform, mithin als „Portale“, die es solchen professionellen Anbietern ermöglichen soll, Personen über Produkte und Dienstleistungen zu informieren, und dabei mithin für den Anbieter zu gewinnen, so folgt unmittelbar, dass ein effektiver und effizienter Zugriff auf die vom Nutzer gesuchte Information zentral ist. Der Nutzer initiiert die Kommunikation mit der Website im Wesentlichen selbst, und kann diesen Dialog, wenn seine Anforderungen nicht erfüllt werden, auch selbst jederzeit abbrechen.

Nutzer auf dem Internet sind heute einer stetig wachsenden Informationsflut ausgesetzt (Lyman & Varian, 2003). Längst haben sich daher Internet-Suchdienste wie Altavista, Excite, und heute Google und Yahoo² als Zugriffshilfe auf Information durchgesetzt und weitgehend klassische Verzeichnisse wenn nicht abgelöst, dann doch ergänzt. Sie ermöglichen den Zugriff quer über das Informationsangebot des Internets mittels Stichwortsuchen. Eine ana-

¹ Gemäss verschiedener Erhebungen, zum Beispiel der WEMF, ist die Internet-Suche der zweithäufigste Nutzungszweck des Internets in der Schweiz (WEMF-Report Plus, 2006).

² Der Internet-Suchdienst Google rangiert konsistent unter den meistbesuchten Websites des Internets. Momentan ist Google die drittmeistbesuchte (Alexa.com) resp. viertmeistbesuchte (Nielsen/Netratings) Site weltweit. Yahoo, die meistbesuchte Site weltweit, bietet ebenfalls sehr prominente Suchfunktionen an.

loge Funktionalität wird daher heute von der Mehrzahl der Nutzer von Unternehmens-Websites erwartet. Dabei ist es aber wichtig, die Suche in Websites von Unternehmen deutlich von der Suche in breiten Internet-Suchdiensten zu unterscheiden. Während letztere einen „breiten“ Informationszugang zu einer sehr grossen Anzahl Websites, unabhängig der einzelnen Inhalte, anbieten (oft als „horizontale Suche“ bezeichnet), bewegen sich die in dieser Studie betrachteten, individuellen Unternehmens-Websites in jedem Fall in einem klar abgegrenzten Gebiet, in welchem gezielt Information zur Verfügung gestellt wird (auch „vertikale Suche“ genannt). Während also Internet-Suchdienste zwischen unzähligen Anbietern gleichartiger oder widersprüchlicher Information gewichten müssen, liegt die Herausforderung für Site-Suchen anders: die Nutzer sind möglichst ohne Umweg auf die meist sehr spezifische gewünschte Information zu führen.

1.2. Die Studie

Die Anforderungen an die Suchfunktion von Unternehmens-Websites sind grundlegend anders als die Anforderungen an Internet-Suchdienste. Die Studie erfasst die Suchfunktion einer Unternehmens-Website („Site-Suche“) als Kombination von Suchapplikation, Konfiguration und dem bereitgestellten Informationsangebot. Die Studie stellt dabei gezielt die Bedürfnisse der Nutzer, nicht einzelne technische Funktionen, in den Vordergrund und betrachtet dabei losgelöst von der eigentlichen Implementierung die Suchfunktion.

Es ist nicht ohne weiteres gesichert, dass Methoden für das öffentlich stark wahrgenommene Anwendungsgebiet der Internet-Suchdienste sich auch bei der Suche auf einer Unternehmens-Website bewähren. Man spricht davon, dass Nutzer von Internet-Suchdiensten oft rein präzisionsorientiert vorgehen: aus der Liste der Suchresultate werden oft nur einige wenige Treffer näher begutachtet (siehe z.B. Schmidt-Mänz & Bomhardt, 2005). Die hohe Redundanz von Information auf dem weiteren Internet ermöglicht es Internet-Suchdiensten relativ grobe Suchmethodiken anzuwenden. Die Optimierung dieser Systeme kann und muss vor allem hinsichtlich anderer Kriterien als der reinen Sucheffektivität erfolgen (Umfang des Informationsangebots, Geschwindigkeit, etc.).

Die Arbeit von (Alvarado et al., 2003) hat gezeigt, dass Nutzer sich nicht mit reiner Stichwortsuche zufrieden geben. Oft wird die Stichwortsuche als ein „Kopfsprung“ ins Informationsangebot aufgefasst, bei dem Gefahr besteht, überzuspezifizieren. Die Nutzer erwarten also Möglichkeiten, Stichwortsuche mit kontrolliertem Navigieren zu kombinieren. Unternehmens-Websites können gezielt diesbezüglich optimiert werden, indem das Informationsangebot für den Zugriff geeignet strukturiert wird.

Die Studie erschliesst die Charakteristiken der Suchfunktionen der untersuchten Websites mit Hilfe von Retrievaltests. Im Rahmen der vorliegenden Studie wird ein wissenschaftlich

fundiertes Evaluationsraster entwickelt, welches eine Reihe klar definierter Tests umfasst, die eine Bewertung der unterschiedlichen Websites ermöglicht. Vier Hauptkriterien werden definiert, nach denen die Site-Suche bewertet wird:

- Der Suchindex
- Der Vergleich zwischen Anfragen und Dokumenten
- Die Benutzerinteraktion
- Die Suchergebnisse

Es ist das Ziel der Studie, den State-of-the-Art in Sachen Site-Suche für Unternehmen zu ermitteln. Dies erlaubt eine Klassierung in gute, mittlere, und schlechte Site-Suchen und identifiziert insbesondere die nötigen Ansatzpunkte, um die Suchfunktion zu verbessern.

1.3. Abgrenzung

Die Studie stellt die Site-Suche als Kombination von Software, Konfiguration und Informationsangebot ins Zentrum. Die Suchfunktion ist daher mithin als Retrievalapplikation zu begreifen, die eine kommunikative Aufgabe für den Betreiber erfüllt; dies steht im Gegensatz zu Internet-Suchdiensten, die auf die technische Aufgabe, mittels Stichwörtern Dokumente zu lokalisieren, reduziert sind.

Es folgt, dass die Studie keine Produktvergleiche anstrebt, da die zur Realisierung der Suchfunktion verwendete Software nur mittelbar am gesamten Suchprozess beteiligt ist und in Verbund mit dem eigentlichen Informationsangebot, sowie der Aufbereitung für die Unternehmens-Website, nur einen Faktor für eine gute Site-Suche darstellt. Tatsächlich ist davon auszugehen, dass etliche der untersuchten Websites auf der gleichen Softwarebasis realisiert sind. Aus diesem Grund sind die gemessenen Leistungswerte z.B. auch nicht auf die Intranetsysteme der gleichen Informationsanbieter übertragbar.

Für die Studie wurden keine Listen von Funktionalitäten zusammengestellt oder Datenblätter der den Site-Suchen zugrunde liegenden Softwarepakete analysiert. Die Studie konzentriert sich auf die Suchfunktionen als Retrievalapplikation, und analysiert daher wie die einer solchen Retrievalapplikation zugrunde liegenden Bedürfnisse gedeckt werden, unabhängig von der konkreten Umsetzung in Systemen.

Die Studie untersucht nicht Suchdienste, die „losgelöst“ vom eigentlichen Informationsangebot eine Vielzahl von unterschiedlichen Websites erschliessen. Die Anforderungen an solche Suchdienste, wie Google, Yahoo und andere, sind deutlich verschieden. Es kann aber nicht

vernachlässigt werden, dass viele potentielle Nutzer einer Site-Suche aus verschiedenen Gründen zu einem Internet-Suchdienst greifen, um eine bestimmte Website zu durchsuchen. Gründe können eine Unzufriedenheit mit den auf der Website angebotenen Suchmöglichkeiten, aber auch eine Gewöhnung an das allgemeinere System sein. Der Anbieter der Information verliert in diesem Szenario sowohl die Kontrolle über die Kommunikation mit dem Suchenden, als auch über die Vollständigkeit und Aktualität des Informationsangebots.

Die Studie strebt keine Rangliste an, noch werden Awards vergeben. Neben dem Wunsch einer Anonymisierung der Resultate gibt es auch fundamentale, wissenschaftliche Gründe, welche ein solches Ranking verbieten. So lässt sich zeigen, dass absolute Leistungswerte aus Retrievaltests, wie sie im Rahmen dieser Studie durchgeführt wurden, relativ grosser Differenzen bedürfen, bis ein statistisch signifikanter Unterschied belegt wird. Der Vergleich von aggregierten Werten und die Klassierung in Leistungsklassen sind hingegen weit stabiler.

1.4. Das Retrievalproblem

Suchfunktionen von Unternehmens-Websites müssen Aufgaben übernehmen, die zu einer klassischen Datenbanksuche fundamental verschieden sind. Datenbanken dienen zur dauerhaften Ablage wichtiger Daten, auf welche unter Nutzung der wohldefinierten Struktur der Ablage wieder zugegriffen werden kann. Dabei kann im Wesentlichen unter Verwendung von Bedingungen gesucht werden: alle Datensätze, welche die Bedingung erfüllen, werden weiterverarbeitet.

Unternehmens-Websites werden von Nutzern konsultiert, welche grundlegend andere Informationsbedürfnisse haben. Die Nutzer suchen auf ihnen „fremden“, unstrukturierten Daten. Die Interpretation der Daten, welche aufgefunden werden, mithin also die Information, welche der Nutzer erhält, differiert von Nutzer zu Nutzer, da sie von Hintergrundwissen, Anwendungszweck und persönlichen Präferenzen subjektiv beeinflusst ist. Der wohl grundlegendste Unterschied ist, dass der Nutzer sich üblicherweise noch nicht schlüssig ist, was überhaupt zu suchen ist: das Informationsbedürfnis ist vage, und reflektiert die Tatsache, dass der Nutzer sich erst mit dem Informationsangebot auseinandersetzen muss, bevor er in der Lage ist, eine gute Formulierung für sein Problem zu finden.

Jegliche Resultate zu einer Suchanfrage müssen daher unter Berücksichtigung dieser Unschärfe zwischen dem Informationsbedürfnis und seiner Formulierung errechnet werden. Da die Suche auf Daten erfolgt, die nicht hinsichtlich der Informationsbedürfnisse der Nutzer vorstrukturiert sind, muss die Formulierung des Informationsbedürfnisses mit verschiedenen Darstellungsmöglichkeiten der gesuchten Information (z.B. verschiedene Umschreibungen des gleichen Sachverhalts) verglichen werden. Die Subjektivität der Wahrnehmung der

gefundenen Information in Hinsicht auf das Informationsbedürfnis (wegen Hintergrundwissen, Präferenzen etc.) führt schliesslich dazu, dass es kein „einzig korrektes“ Suchresultat gibt.

Man spricht in der Literatur vom „Retrievalproblem“: „Das Auffinden von möglichst viel relevanter Information bei gleichzeitigem Minimieren der ebenfalls gelieferten irrelevanten Information“ (Van Rijsbergen, 1979).

1.5. Relevanz

Wie angedeutet, hängt die Beurteilung der Relevanz einer Information (z.B. eines Dokuments) hinsichtlich eines Informationsbedürfnisses von vielen Faktoren ab:

- Vor-/Hintergrundwissen
- Reihenfolge des Auffindens
- Wandelnde Informationsbedürfnisse
- Persönliche Präferenzen
- Vollständigkeit der Antwort

Dies erschwert eine absolute Bewertung von Suchresultaten in wissenschaftlich, objektiven Sinn. Zur Bewertung der Effektivität von Retrievalsystemen hat sich das sogenannte Cranfield-Paradigma durchgesetzt (Cleverdon, 1967). Dabei wird grossen Wert auf den Vergleich zwischen Systemen gelegt. Die Erkenntnisse aus der Evaluation sind allgemein desto aussagekräftiger, desto mehr Systeme in die Evaluation einbezogen worden sind (siehe auch Voorhees, 2002). Die vorliegende Studie versucht diesem Umstand insofern Rechnung zu tragen, als dass 56 verschiedene Websites evaluiert wurden. Dies, gekoppelt mit einer grossen Anzahl Einzeltests zu den einzelnen untersuchten Kriterien, erlaubt es, zuverlässige Aussagen, insbesondere hinsichtlich der typischen Performance von Site-Suchen, zu machen.

Die Studie tritt also in die Fussstapfen grosser Retrievalsystemevaluations wie TREC³ (Harman, 1995), CLEF⁴ (Peters & Braschler, 2001), und NTCIR⁵ (Kando,2004), bleibt aber im Gegensatz zu diesen breiter abgestellten Unternehmungen fest dem Grundsatz verpflichtet, konkrete Unternehmens-Websites, nicht unterliegende Systeme, von Anbietern aus der Schweiz zu erfassen.

³ <http://trec.nist.gov>

⁴ <http://www.clef-campaign.org>

⁵ <http://research.nii.ac.jp/ntcir/>

2. Testanordnung und Durchführung

2.1. *Evaluations-Raster*

Für die Untersuchung der Sitesuche haben Eurospider, HTW Chur, namics und Zürcher Hochschule Winterthur gemeinsam ein Evaluationsraster entwickelt. Das Raster ist in vier Bereiche gegliedert (Suchindex, Vergleich von Anfragen und Dokumenten, Benutzerinteraktion und Suchergebnisse), mit deren Hilfe die wesentlichen Dimensionen einer Sitesuche abgedeckt werden. Diese Hauptkriterien werden jeweils durch mehrere Subkriterien gebildet, und jedes Subkriterium wiederum durch eine Reihe von Einzeltests ermittelt.

So besteht beispielsweise das Hauptkriterium Anfrage- und Dokumentenvergleich aus den drei Kriterien Anfrageausführung, Ausdrucksstärke der Abfragesprache und Metadatenqualität. Die Anfrageausführung wird durch drei Einzeltests ermittelt:

- Werden sehr lange Anfragen ausgewertet?
- Können Phrasen mit Stoppwörtern gesucht werden?
- Ist die Suche gegenüber Flexionen (Wortbeugungen) robust?

Diese Strukturierung ermöglicht, einzelne Aspekte der Sitesuche differenziert zu bewerten und zu würdigen.

Das Evaluationsraster besteht überwiegend aus manuellen sowie einigen ausgewählten automatische Tests. Für jeden einzelnen Test wurde ein Testprotokoll formuliert, welches Punkt für Punkt durchgegangen werden musste. Für die Codierung der Ergebnisse wurden ebenfalls feste Vorgaben gemacht.

Das vollständige Evaluationsraster wurde von Eurospider, HTW Chur und namics in einem Pre-Test zunächst im Hinblick auf seine Durchführbarkeit hin überprüft. Zu diesem Zweck wurden die Site-Suchen von drei grossen Schweizer Versicherungsgesellschaften der Evaluation unterzogen. Die Pre-Tests führten unter anderem zum Ausschluss verschiedener nicht durchführbarer Tests. Auf Tests, die nicht isoliert, ohne Intervention der Betreiber stattfinden konnten, wurde verzichtet. Stattdessen wurden zum Teil neue Tests mit einbezogen. Ebenfalls wurde die Reproduzierbarkeit der Tests bzw. der Testanweisungen überprüft und letztere gegebenenfalls präzisiert. So konnte der subjektive Eindruck des Testers bestmöglich eliminiert werden.

Nach dem Pre-Test erfolgte die eigentliche Evaluation der insgesamt 56 getesteten Websites. Die manuellen Tests wurden durch zwei Studenten der Zürcher Hochschule Winterthur

sowie eine namics-Mitarbeiterin durchgeführt. Zu diesem Zweck erhielten sie eine 36-seitige Testanleitung. Die automatischen Tests wurden durch Eurospider standardisiert ausgeführt. Der Gesamtaufwand für die Evaluation einer Site-Suche lag bei 4-6 Stunden.

Da die Tests „von aussen“, also über das öffentliche Internet durchgeführt werden konnten, wurden die Betreiber der getesteten Sites über den Testzeitpunkt nicht informiert. So konnten vorgängige Anpassungen an den getesteten Systemen ausgeschlossen werden.

Im Hinblick auf die Auswertung der einzelnen Assessments wurden die Einzeltests, Subkriterien und Hauptkriterien gewichtet. Dies erfolgte nach Einschätzung der beteiligten Partner. Die Gewichtung wird erleichtert durch den Aufbau der Studie mit weitgehend unabhängigen Haupt- und Subkriterien. Die normalisierten Werte wurden dann aggregiert und zur weiteren Analyse graphisch aufbereitet.

2.2. *Untersuchte Websites:*

| | |
|-----------------------|--------------------|
| www.abb.ch | www.national.ch |
| www.admin.ch | www.novartis.ch |
| www.allianz-suisse.ch | www.pax.ch |
| www.bankcoop.ch | www.phonak.ch |
| www.baselland.ch | www.raiffeisen.ch |
| www.bcv.ch | www.railaway.ch |
| www.be.ch | www.ringier.ch |
| www.bekb.ch | www.roche.com |
| www.bfs.admin.ch | www.sgkb.ch |
| www.bkb.ch | www.sunrise.ch |
| www.blkb.ch | www.suva.ch |
| www.bs.ch | www.swiss.com |
| www.cablecom.ch | www.swisslife.ch |
| www.converium.com | www.swissre.com |
| www.credit-suisse.com | www.synthes.com |
| www.css.ch | www.tkb.ch |
| www.ethzh.ch | www.ubs.com |
| www.generali.ch | www.unibe.ch |
| www.gkb.ch | www.unique.ch |
| www.helvetiapatria.ch | www.valiantbank.ch |
| www.holcim.ch | www.vbs.admin.ch |
| www.juliusbaer.ch | www.visana.ch |
| www.kaba.ch | www.zh.ch |
| www.lonza.ch | www.zhwin.ch |
| www.lukb.ch/ | www.zkb.ch |
| www.migrosbank.ch | www.zug.ch |
| www.mobi.ch | www.zugerkb.ch |
| www.nab.ch | |

Zusätzlich wurde die Website eines weiteren Schweizer Grossunternehmens, das nicht genannt werden möchte, der Evaluation unterzogen.

3. Gesamtergebnis

3.1. Boxplots

Die Studie präsentiert die Resultate in Form von aggregierten Resultatwerten, nicht von Einzelwerten. Dies begründet sich, wie in der Einleitung ausgeführt, in der Interpretierbarkeit der Resultate: die absoluten Einzelwerte sind statistisch weit weniger aussagekräftig als relative Vergleiche und aggregierte Durchschnittswerte.

Zur Diskussion ist die Darstellungsform des Boxplots besonders geeignet. Sie liefert in einer kompakten grafischen Darstellung einen Überblick über mehrere Charakteristiken der Gesamtmenge der Resultate aller Probanden.

Der Boxplot stellt dar:

- die mittlere Leistung (repräsentiert durch den Median⁶). Dargestellt durch eine breite horizontale Linie in der Box
- die Streuung der individuellen Einzelresultate (dargestellt durch eine breite Box). Die Box umfasst 50% aller getesteten Probanden, sie reicht vom 1. Quartil⁷ bis zum 3. Quartil⁸
- die Schiefe der Verteilung der Einzelresultate (dargestellt durch die Lage des Medians innerhalb der Box)
- die Minimalleistung und Maximalleistung (dargestellt durch zwei „Antennen“, die von der Box in beiden Richtungen je die 25% schlechtesten resp. besten Resultate umfassen.
- Als Ausreisser werden Werte bezeichnet, die mehr als 1.5-mal den Interquartilabstand (= Höhe der Box) von der Box entfernt sind. Die Ausreisser werden in den folgenden Grafiken nicht ausgewiesen.

⁶ Je 50% aller Site-Suchen haben Leistungen unter resp. oberhalb des Medians erzielt.

⁷ Auch „unteres Quartil“ und „0,25-Quantil“ genannt. Ein Zahlenwert, der eine Leistung repräsentiert, die von 75% der Probanden übertroffen wurde.

⁸ Auch „oberes Quartil“ und „0,75-Quantil“ genannt. Ein Zahlenwert, der eine Leistung repräsentiert, die von 25% der Probanden übertroffen wurde.

3.2. *Diskussion der Hauptkriterien*

Wir beginnen die Diskussion der Resultate im Folgenden mit einer Betrachtung der vier Hauptkriterien:

- Der Suchindex
- Der Vergleich von Anfragen und Dokumenten
- Die Benutzerinteraktion
- Die Suchergebnisse

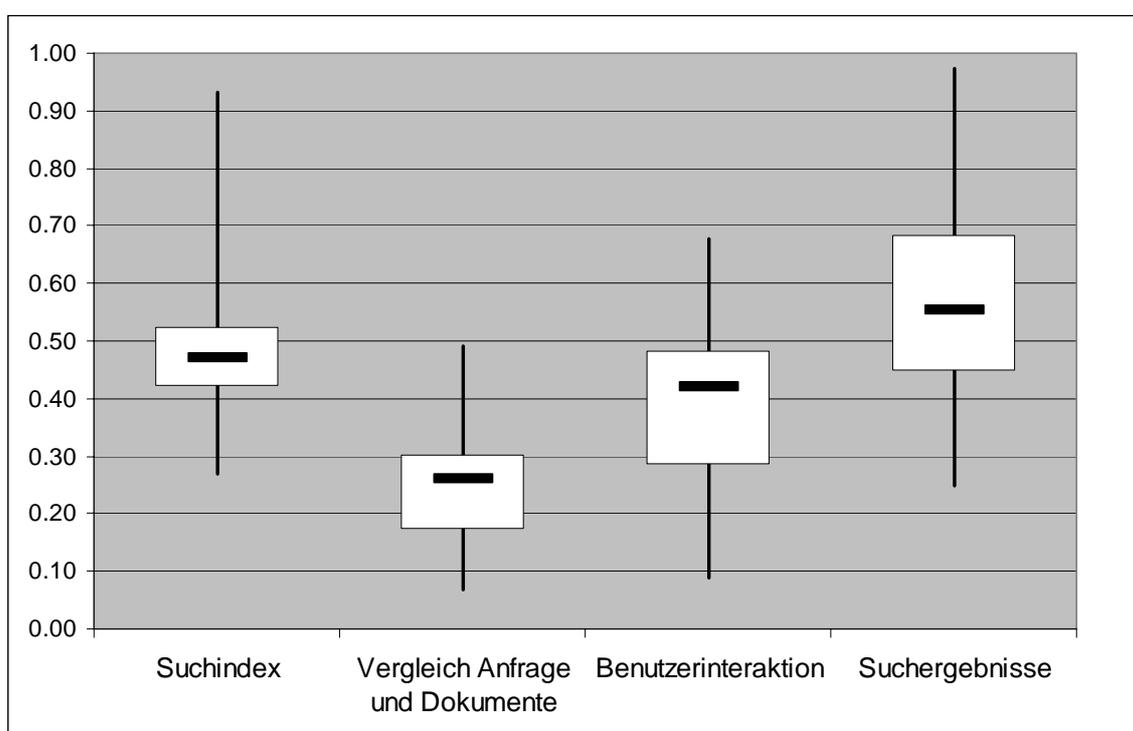


Abbildung 2: Boxplot der vier Hauptkriterien

Es fällt bei Betrachtung des Boxplots (siehe Abbildung 2) auf, dass die Kennwerte der vier Hauptkriterien deutlich verschieden sind. Während absolute Leistungswerte grundsätzlich vorsichtig einzuordnen sind, da sie massgeblich durch die Ausgestaltung der einzelnen Einzeltests beeinflusst sind, ist trotzdem augenfällig, dass die Leistungswerte im Bereich „Vergleich von Anfragen und Dokumenten“ durchs Band gegenüber den anderen Kriterien abfallen. Die mittlere Leistung ist bei 0.26 angesiedelt, während sie für die anderen drei Kriterien deutlich die Marke von 0.4 übertrifft. Die Streuung ist dabei am zweitgeringsten, was heisst, dass die Leistungswerte der grossen Mehrheit der Site-Suchen in diesem Bereich als ungenügend betrachtet werden muss. Nicht zuletzt ist auffällig, dass auch die Maximalleistung, d.h. die beste getestete Unternehmens-Website, deutlich unter der theoretisch möglichen

Bestleistung zurückbleibt. Dabei erreichen nur gerade 3 weitere Site-Suchen mindestens 80% der Leistung dieses „Klassenbesten“.

| Statistiken | Suchindex | Vergleich Anfragen/- Dokumente | Benutzer- interaktion | Suchergebnisse |
|-------------|-----------|-----------------------------------|--------------------------|----------------|
| Maximum | 0.93 | 0.49 | 0.68 | 0.97 |
| 3. Quartil | 0.52 | 0.30 | 0.48 | 0.68 |
| Median | 0.47 | 0.26 | 0.42 | 0.55 |
| 1. Quartil | 0.42 | 0.17 | 0.28 | 0.45 |
| Minimum | 0.27 | 0.07 | 0.09 | 0.25 |

Die restlichen drei Hauptkriterien weisen eine mittlere Leistung auf, die sich in einem Band um 0.5 befindet. Gewichtige Unterschiede kommen aber zu Tage, wenn die Streuung betrachtet wird: bei weitem die geringste Streuung weist das Hauptkriterium „Suchindex“ auf, wo sich die Mehrzahl der Site-Suchen nahe an die mittlere Leistung anlehnt. Dabei ist der Abstand zum ersten resp. dritten Quartil bemerkenswert symmetrisch, d.h., der Median kommt in der Mitte der Box zu liegen. Anders sieht die Lage an den beiden Enden der Leistungsskala für das Hauptkriterium „Suchindex“ aus: der Vorsprung der besten Site-Suchen zur mittleren Leistung ist mehr als doppelt so gross wie der Rückstand der schwächsten Websites. Vier Site-Suchen bringen eine Leistung, die im statistischen Sinne als „Ausreisser“ zu werten ist⁹. Die Site-Suche mit der Bestleistung ist dabei im statistischen Sinne gar als „extremer Ausreisser“ zu werten¹⁰. Die Bestleistung ist dabei sehr nahe am theoretischen Maximum (0.93).

Ein ähnlicher, sogar leicht höherer Bestwert ist für das Hauptkriterium „Suchergebnisse“ zu verzeichnen. Allerdings ist für dieses Kriterium die Streuung deutlich grösser als für das Kriterium „Suchindex“. Die Bestwerte sind deshalb im statistischen Sinne nicht als Ausreisser aufzufassen. Wieder ist der Vorsprung der besten Site-Suchen auf die mittlere Leistung grösser als der Rückstand der Nachzügler, wobei dieser Unterschied aber weniger ausgeprägt ist. Die Bandbreite über alle Websites ist für kein Hauptkriterium so gross wie für die Suchergebnisse.

⁹ D.h., ihre Leistungswerte sind mehr als 1.5-mal den Interquartilsabstand (die Höhe der Box) von der Box entfernt.

¹⁰ D.h., der Leistungswert ist mehr als 3-mal den Interquartilsabstand von der Box entfernt.

Bleibt die Betrachtung des vierten und letzten Hauptkriteriums, der „Benutzerinteraktion“. Die mittlere Leistung ist im Groben vergleichbar mit der für das Kriterium „Suchindex“ gemessenen, die Streuung ist aber ungleich höher, wobei dies für dieses Kriterium vor allem nach unten gilt: der Rückstand der Nachzügler ist hier höher als bei den anderen Kriterien, und mehr Websites hinken der mittleren Leistung hinterher.

3.3. Gesamthafte Betrachtung

Es ist weiterhin festzuhalten, dass die vier Bestleistungen in den vier Hauptkriterien von vier unterschiedlichen Websites erreicht wurden. Das zeigt, dass ein grosser Abstand nicht nur zur theoretischen Bestleistung besteht, sondern auch zu einer real heute erreichbaren Leistung¹¹. Düster sieht es am anderen Ende der Skala aus: Mehr als die Hälfte der Websites konnte sich für kein Hauptkriterium in der Spitzengruppe etablieren. Hier ist davon auszugehen, dass die betreffenden Informationsanbieter grosse Probleme haben werden, mit der vorliegenden Site-Suche die Kommunikation mit den Nutzern erfolgreich zu gestalten.

Betrachtet man die Hauptkriterien paarweise, so ist ein ähnliches Muster zu erkennen. Für kein Paar von Hauptkriterien gibt es überzeugende Anzeichen, dass dieselben Websites konsistent die besten Ergebnisse erreichen. Dies ist einerseits ein Hinweis auf die weitgehende Orthogonalität der Hauptkriterien, wie während der Entwicklung des Evaluationsrasters angestrebt, aber andererseits auch Anzeichen für das potentiell grosse Potential für Verbesserungen der real existierenden Websites. So sind nur drei Websites zu finden, die für drei der vier Kriterien in der „Spitzengruppe“ in Sachen Leistung rangieren (d.h., unter den 10 besten Testresultaten für das jeweilige Hauptkriterium). Keine Website kann sich für alle vier Hauptkriterien gleichmässig gut auszeichnen. Auf der anderen Seite ist nur ein System für mindestens drei der vier Hauptkriterien in der Gruppe der schwächsten Websites anzutreffen.

3.4. Allgemeiner Eindruck

Die Tester hatten im Rahmen der Evaluation ihre persönlichen Eindrücke gesondert als „allgemeiner Eindruck“ festzuhalten. Diese Angaben wurden separat ausgewertet. Die Tester wurden gebeten, mittels einer Note zwischen 1 („ungenügend“) und 10 („exzellent“) den „Gesamteindruck“, welchen sie von der Suchfunktion gewonnen haben, zu bewerten.

¹¹ Unter der Voraussetzung, dass man die Hauptkriterien als weitgehend unabhängig betrachtet. In der Realität sind natürlich allfällige Zielkonflikte zwischen den einzelnen Kriterien nicht völlig auszuschliessen

Die durchschnittliche Bewertung über alle Websites liegt bei 4.84 (Median 5), und damit analog zu den einzelnen Hauptkriterien weit vom theoretischen Maximum entfernt. Das Erlebnis der Tester bei der Nutzung der Site-Suchen deckt sich also durchaus mit der objektiven Auswertung der Einzeltests.

Die Tester machten des Weiteren rege von der Möglichkeit Gebrauch, einen kurzen Kommentar zum Umgang mit der Suchfunktion zu geben. Notiert wurden vor allem Probleme, welche aufgetreten waren. So vermissten die Tester immer wieder „griffige“ Einschränkungsmöglichkeiten, die es erlauben würden, die Suche geeignet zu lenken. Es wurde ferner vermerkt, dass oft Operatoren bei der Suche fehlen oder hart verdrahtet sind (z.B. hart verdrahtete „UND“-Verknüpfung der Suchbegriffe: es können keine Dokumente gefunden werden, die nur einen Teil der Suchbegriffe beinhalten). Auf die Suche konnte nicht im erwünschten Mass Einfluss genommen werden. Das Fehlen eines Phrasenoperators wurde bei verschiedenen Websites beanstandet. Mehrmals wurde einem Erstaunen Ausdruck verliehen, dass die Site-Suche eines renommierten Unternehmen so „simpel“ ausgeführt ist und dass leistungsfähige Funktionen fehlen.

Neben dieser Bewertung des Gesamteindrucks wurde den Testern die Möglichkeit gegeben, einen „Spasfaktor“ zu bewerten, der ausdrücken sollte, ob der Umgang mit der Suchfunktion eher als angenehm oder belastend wahrgenommen wurde. Dabei wurde eine Skala mit drei Werten verwendet (0 = kein Spass, 1 = war in Ordnung, 2 = hat Spass gemacht). Die durchschnittliche Bewertung lag bei 0.82 Punkten (Median: 1). Als negativ wurden speziell Fehleranfälligkeit, langsame Suchvorgänge, mangelnde Transparenz und unerwartetes Verhalten genannt. Beliebt bei den Testern waren Site-Suchen, die einfach zu bedienen waren, schnell auf Eingaben reagierten und mit einer ausgefeilten Präsentation zu überzeugen wussten.

4. Suchindex

Das Kriterium Suchindex beurteilt die Qualität der Datengrundlage einer Site-Suche. Diese technisch für den Zugriff optimierte Datengrundlage, der so genannte Index¹², beeinflusst Möglichkeiten, definiert aber auch Einschränkungen sowie qualitative Aspekte der darauf agierenden Suchapplikation. Plakativ gesagt kann die Suchanwendung nur verarbeiten, was im Index vorhanden ist. Fehlt ein Dokument, wird es nie auf der Suchtrefferliste erscheinen oder ist ein Dokument fehlerhaft im Index vorhanden, so wird die Suchanwendung nie ein korrektes Resultat liefern können. Die Qualität des Index wurde aufgrund folgender Subkriterien beurteilt: Vollständigkeit, Aktualität und Anfrage- und Dokumentenerschliessung.

Die Medianperformance (vgl. Abbildung 2) der Bewertungen des Kriteriums Suchindex liegt mit 0.47 im selben Bereich wie zwei der anderen drei Hauptkriterien. Das Resultat ist eher ernüchternd, lässt sich in der Folge aber gut erklären. Die geringe Streuung der Hälfte alle Probanden und die Mittigkeit des Medians weisen darauf hin, dass ein wesentlicher Teil der evaluierten Anwendungen eine ähnliche Strategie bei der Implementierung verfolgt. Auffällig ist zudem die Differenz der Extrema, welche aufgrund der Unterschiedlichkeit der eingesetzten Suchtechnologie jedoch verständlich ist und daher kaum vertieft werden muss.

| Statistiken | Vollständigkeit | Aktualität | Anfrage- und Dokumentenerschliessung |
|-------------|-----------------|------------|--------------------------------------|
| Maximum | 1.00 | 1.00 | 1.00 |
| 3. Quartil | 0.58 | 0.00 | 1.00 |
| Median | 0.44 | 0.00 | 1.00 |
| 1. Quartil | 0.33 | 0.00 | 0.83 |
| Minimum | 0.11 | 0.00 | 0.50 |

Die Bewertung der Subkriterien ist wie folgt (siehe Abbildung 3):

¹² Vgl. Witten et al. (1999) und Frakes & Baeza-Yates (1992)

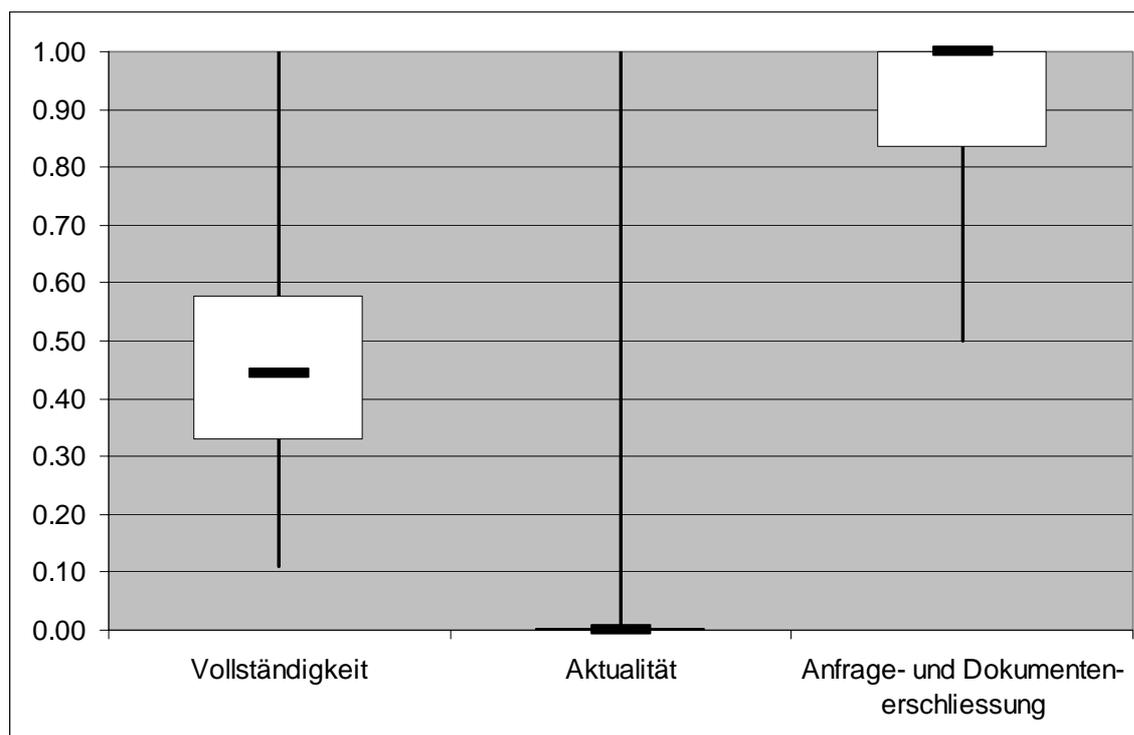


Abbildung 3: Hauptkriterium “Suchindex“: Boxplot der Subkriterien

Bei der **Vollständigkeit des Suchindex**es wurde geprüft, ob die Erschliessung der geprüften Website durch die Suchfunktion möglichst vollständig ist. Als Referenzsysteme zur Beurteilung von Stichproben dienten öffentliche Suchdienste wie Google und Yahoo. Des Weiteren wurden – insofern auf der Site vorhanden – lange, respektive komplexe URLs¹³ und binäre Dokumente in der Form von Adobe PDF-Dateien¹⁴ auf deren Indexstatus getestet. Der Median von 0.44 ist insofern bemerkenswert, da die Anforderung alle öffentlich zugänglichen Seiten des Webangebotes über die eigene Suchfunktion zugänglich zu machen, von jedem Anbieter angestrebt wird. In diesem Sinne wird also eine Minimalanforderung nur ungenügend erfüllt. Zudem zeigt die maximale Leistung von 1.00, dass die Aufgabe lösbar ist. Schlussendlich sind alle benötigten Elemente im Wirkungsbereich des Betreibers. Ein aufgetretenes Problem ist der Umgang mit PDF-Dateien. Dabei ist vor allem der Umstand von Bedeutung, dass es viele verschiedene Typen und Konfigurationen von PDFs gibt, welche sich bei der Extraktion von Text, so wie von der Suchfunktion benötigt, unterschied-

¹³ Uniform Resource Locators (URL) ist ein Adressierungsschema für Webseiten. Dieses wird insbesondere bei der Verlinkung angewendet.

¹⁴ Portable Document Format (PDF) ist eine von der Firma Adobe Systems entwickelte Familie von Dateiformaten, mit welchen sich graphisch gestaltete 2D oder 3D Dokumente plattform- und (je nach Inhalt) auflösungsunabhängig austauschen lassen.

lich verhalten. Interessant ist des Weiteren, dass alle Testpersonen zusätzlich zu den strukturiert abgefragten Kriterien regelmässig ein Augenmerk auf die Funktionalität der Suchfunktionen hinsichtlich PDFs legten und die Sites im Bezug darauf lobten, oder eben kritisierten.

Der auffälligste Wert ist die **Aktualität des Suchindexes**, bei welchem das Gros aller Probanden den Test nicht bestanden. Anforderung hierbei war es, dass bei inhaltlichen Änderungen auf der Website, der Index und somit die Suchfunktion diese sofort berücksichtigen. Zwei typische Schwächen waren einerseits neu angelegte oder geänderte Seiten, beispielsweise eine aktuelle Medienmitteilung welche mittels der Suchfunktion nicht aufgefunden werden konnte. Andererseits der Fall, dass in der Trefferliste Links zu Zieldokumenten aufgeführt wurden, welche nicht mehr existierten. Bei einem Klick auf den Link wurde der User dann mit einer Fehlermeldung konfrontiert. Grund für diese Fehlerklasse ist eine mangelhafte Integration der Suchtechnologie mit dem Pflegeprozess der Inhalte. Der Autor oder spätestens das Content Management System sind über die Änderung einer Seite informiert – weshalb aber nicht die Suchfunktion?

Die **Anfrage- und Dokumentenerschliessung** schliesst mit einem Median von 1 und einem geringen Quartilsabstand gut ab. Sehr gut bestanden die getesteten Websites die Anforderung diakritische Zeichen wie ä oder à mittels Normalisierung korrekt zu behandeln. Dazu beigetragen hat sicherlich die Sprachvielfalt in der Schweiz, wo die Anforderung häufig auftritt. Weniger einheitlich erfüllt wurde die Anforderung an eine korrekte Tokenisierung, also die Behandlung von Sonderzeichen innerhalb von Suchbegriffen wie „E-Mail“ oder der Dezimaltrenner ab CHF 1'000. Bei der Detailbewertungen zeigt sich zudem, dass Suchfunktionen, welche diakritische Zeichen nicht beherrschen, auch bei Sonderzeichen scheitern.

Die Hypothese ist zulässig, dass die Aufgabe der Suchfunktion und deren Wichtigkeit von den Personen, welche die technische Umsetzung und den Betrieb leisten, kaum respektiert wird. Selbst grundlegende Anforderung an die Datenhaltung der Suchapplikation wird in Bereichen vernachlässigt, welche einfach beherrschbar und einfach testbar sind: Aktualität und Vollständigkeit. Hauptgrund, zusätzlich zum geringen Augenmerk, ist vermutlich die passive Erschliessung der Daten über ein Crawling, das nicht nach Bedarf, aber zeit- oder lastgesteuert ist. Dies anstelle einer ereignisgesteuerten Aktualisierung.

5. Anfragen und Dokumente vergleichen

Ein Suchresultat wird berechnet, indem die Anfrage mit den erschlossenen Dokumenten verglichen wird. Das Vergleichen von Anfragen und Dokumenten wurde aufgrund folgender Subkriterien beurteilt: Auswertung der Anfrage, Ausdrucksstärke der Abfragesprache und Metadatenqualität.

Beim Vergleichen von Anfragen und Dokumenten geht es nicht primär um die Suchqualität, sondern darum, ob die Auswertung einer Benutzeranfrage technisch sauber realisiert wurde, was natürlich auch einen Einfluss auf die Suchqualität hat. Der Boxplot aus Abschnitt 3.2 (siehe Abbildung 2) zeigt schlechte Resultate (50% kommen nicht über 0.26, 75% kommen nicht über 0.3 und niemand kommt über 0.50). Die Boxplots für die Subkriterien zeigen ein differenzierteres Bild (siehe Abbildung 4):

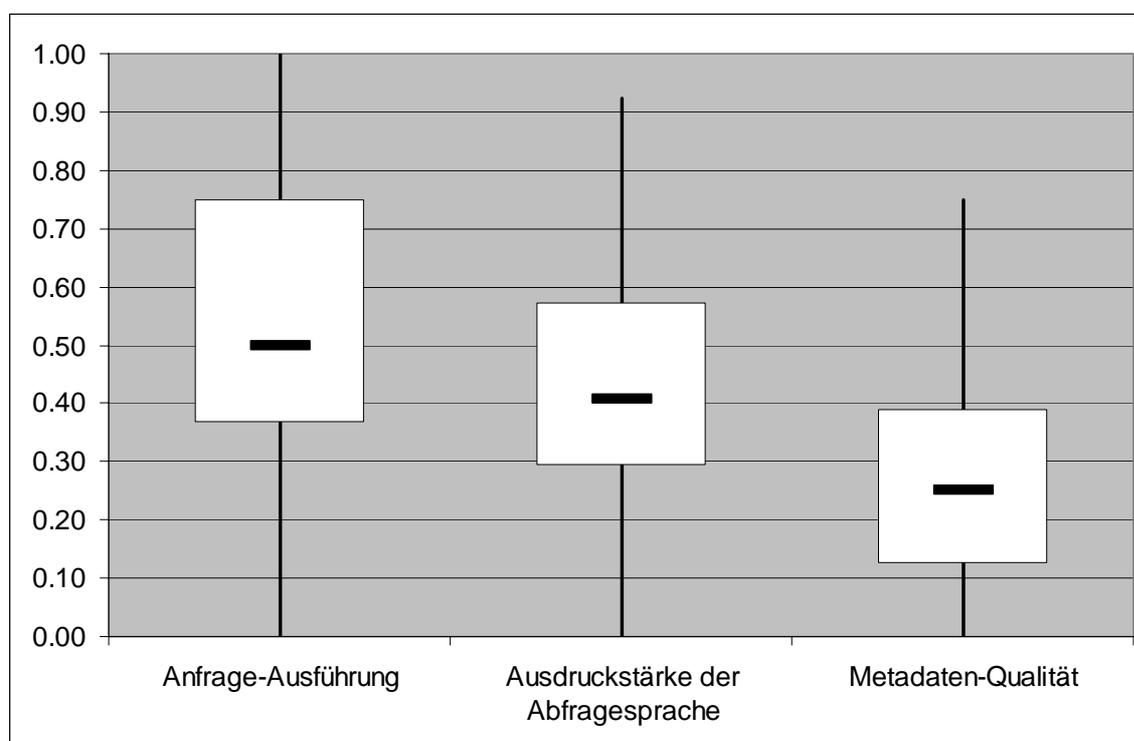


Abbildung 4: Hauptkriterium "Vergleich Anfrage / Dokumente": Boxplot der Subkriterien

| Statistiken | Anfrageausführung | Ausdrucksstärke der Abfragesprache | Metadatenqualität |
|-------------|-------------------|------------------------------------|-------------------|
| Maximum | 1.00 | 0.92 | 0.75 |
| 3. Quartil | 0.75 | 0.57 | 0.39 |
| Median | 0.50 | 0.41 | 0.25 |
| 1. Quartil | 0.36 | 0.29 | 0.16 |
| Minimum | 0.00 | 0.00 | 0.00 |

Die Retrieval-Applikationen schneiden bei der Anfrageausführung am besten und bei der Metadatenqualität am schlechtesten ab. Besonders auffällig ist, dass 75% alle Applikationen eine schlechte Metadatenqualität aufweisen, d.h. Werte zwischen 0.0 und 0.4. Die Maximalwerte zeigen, dass es durchaus möglich ist, beim Vergleichen von Anfragen und Dokumenten gute Werte zu erzielen.

Die Site-Suchen, welche bei der **Anfrageausführung** schlecht abgeschnitten haben, sind eher Datenbank- als Retrieval-Applikationen. Ihnen fehlt die erforderliche Robustheit, um mit beliebigen unstrukturierten Abfragen fertig zu werden. Die beim Retrieval-Problem inhärente Unsicherheit (Abschnitt 1.4) wird bei Datenbankapplikationen nicht berücksichtigt.

Wenn man beispielsweise eine Textpassage in einem wichtigen Dokument in die Zwischenablage kopiert und anschliessend in das Suchfeld einfügt, dann sollte das Dokument gefunden werden. Ein leeres Resultat ist nicht akzeptabel. Das Dokument sollte auch gefunden werden, wenn die Textpassage aus einer alten Version stammt, in der leicht andere Formulierungen (z.B. Plural anstatt Singular) verwendet werden.

Bei der **Ausdrucksstärke der Abfragesprache** fällt die grosse Streuung auf (Werte zwischen 0.00 und 0.92). Die schlechtesten Systeme unterstützen keine Operatoren, so dass in besonderen Fällen das Informationsbedürfnis nicht adäquat ausgedrückt werden kann. Beispiele: Tel AND NOT "Tel Aviv".

Die Beurteilung der **Metadaten** fiel ernüchternd aus und zeigt, dass die Produktion der Online-Inhalte in vielen Fällen noch nicht durchgängig ist. Erstaunlich ist, dass in einigen Fällen nicht einmal ein korrektes Datum angegeben wird. In vielen Fällen fehlt die Angabe, zu welcher Teilkollektion ein gefundenes Dokument gehört, was dann im Suchprozess das explizite Ein- und Ausschliessen von Teilkollektionen erschwert.

Im Folgenden wird die Anfrageausführung anhand einiger Beispiele erläutert:

- Der Hauptsatz aus dem Leitbild wie beispielsweise "Die Unternehmung XY ist die führende Anbieterin von XY-Produkten mit einer internationalen Ausrichtung" sollte nicht zu einem leeren Resultat führen.
- Die bei XY für die Kommunikation verantwortliche Person möchte mit der Anfrage "gegen XY" herausfinden, ob auf der Web Site etwas Nachteiliges über ihre Organisation XY publiziert wird. Dies ist nur möglich, wenn in Phrasen auch Stoppwörter berücksichtigt werden. Stoppworte, wie beispielsweise "gegen" werden bei der Erschliessung häufig ignoriert, weil sie wenig Informationsgehalt haben. Hingegen als Bestandteil von Phrasen können Stoppwörter durchaus wichtig sein.
- Wenn auf einer Website Kühlschränke angeboten werden und ein Kaufinteressent nach Kühlschrank sucht, weil er nur einen Kühlschrank sucht, dann sollte er das Angebot trotzdem finden. Besonders bei deutschsprachigen Websites ist es wichtig, dass die Suchapplikationen auch Flexionen (deklinierte und konjugierte Wortformen) berücksichtigt.
- Ebenfalls wichtig bei deutschsprachigen Inhalten ist die Behandlung von Komposita. Die Anfrage Konzernleitungsmitglied soll selbstverständlich Texte finden, wo erwähnt ist, dass XY seit neuestem Mitglied der Konzernleitung ist.

Im Folgenden wird die Ausdrucksstärke der Abfragesprache anhand einiger Beispiele erläutert:

- Bei einer sauberen Anfrageausführung soll beispielsweise mit der Phrase "4 GB" problemlos Texte gefunden werden, in denen sich die Einheit GB auf die Quantität 4 bezieht, und nicht etwa Laptop Angebote mit einem 15.4" WXGA Display und einer 80 GB Hard-disk.
- Bei mehrdeutigen Suchbegriffen ist es praktisch, die nicht gewünschte Bedeutung mit dem Ausschlussoperator auszuschliessen. Beispielsweise kann Code im Sinne von "Code of Conduct", "Program Code", "Da Vinci Code", etc. gemeint sein. Mit der Anfrage Code –Vinci wird die Bedeutung Da Vinci Code ausgeschlossen.
- Zahlreiche Suchapplikation verknüpfen die Suchbegriffe automatisch mit dem UND-Operator, so dass nur Dokumente gefunden werden, welche alle Suchkriterien erfüllen. In diesem Fall ist es wenig sinnvoll den Einschlussoperator „+“ anzubieten, weil er keinen Einfluss auf das Suchresultat hat.

- Der Einschlussoperator ist hilfreich, wenn man beispielsweise Eigenschaften zu einem Produkt XY sucht, aber nicht genau weiss, wie diese Eigenschaften formuliert sind. In diesem Fall kann man mit der Anfrage „+XY MöglicheEigenschaftA MöglicheEigenschaftB MöglicheEigenschaftC“ erzwingen, dass nur Information zu XY gefunden werden.
- Die in letzter Zeit viel gerühmte Boole'sche Logik ist gemäss seriösen Untersuchungen viel weniger wichtig, als dies uns gewisse Recherchierspezialisten weismachen möchten¹⁵. Boole'sche Operatoren sollten nur in Ausnahmefällen eingesetzt werden, dann müssen sie aber korrekt und in ihrer vollen Mächtigkeit genutzt werden können. Die Boole'sche Anfrage X OR Y AND X deutet beispielsweise darauf hin, dass der Benutzer die Boole'sche Logik nicht verstanden hat.

Im Folgenden wird die Metadatenproblematik anhand einiger Beispiele erläutert:

- Untersuchungen (Dumais et al. 2003) haben gezeigt, dass das Datum beim Wiederfinden von Dokumenten, die man schon einmal gesehen hat, eine wichtige Rolle spielt, vor allem deshalb, weil man das Erscheinen des gesuchten Dokumentes zeitlich einordnen kann (erschien vor bzw. nach ...). Neben der Möglichkeit, gefundene Dokumente chronologisch zu sortieren, erfordert dies aber zwingend ein korrektes Datum in den Metadaten. Dieses Datum muss in einem interpretierbaren Format vorliegen so, dass man es mit anderen Daten vergleichen kann.
- Ein im Internet publiziertes PDF-Dokument sollte mindestens Titel und Verfasser in den Dokumenteneigenschaften korrekt wiedergeben. Nicht selten findet man Dokumente, welche mit "Vorlage XY" betitelt sind. Für eine Suchapplikation ist es fast unmöglich herauszufinden, ob der Titel in den Metadaten (bei Word und PDF: Dokumenteneigenschaften) sorgfältig erstellt oder nur ein Dummy Titel ist. Umgekehrt spielt gemäss Tombros & Sanderson (1998) der Titel eine sehr grosse Rolle, beim Entscheid ob im Suchresultat ein gefundenes Dokument angeklickt wird oder nicht.

¹⁵ In Harman (1993) wird begründet: „Systems that provide an unranked set of documents are known to be less effective and therefore were not tested in the TIPSTER/TREC programs.“ In Fairthorne (1961) schreiben Moers und Calvin: "It is a common fallacy, underwritten at this date by the investment of several million dollars in a variety of retrieval hardware, that the algebra of George Boole (1847) is the appropriate formalism for retrieval system design. This view is as widely and uncritically accepted as it is wrong."

6. Benutzerinteraktion

Das Kriterium Benutzerinteraktion beurteilt den Teil der Suchapplikation, mit welchem der Nutzer direkt arbeitet resp. kommuniziert. Bei allen getesteten Websites handelte es sich dabei um eine einfache Volltextsuche mit Texteingabefeld und Suchbutton sowie zur Resultatsübersicht eine rangierte Trefferliste mit Trefferzitaten. Die meisten Websites ergänzen diese zwei Ansichten um einen funktional ausführlicheren Eingabedialog, meist „erweiterte Suche“ genannt, mit Optionen wie der Einschränkung der Kollektion. Sehr wenige Websites bieten eine aktivere Nutzerführung wie Anfrageerweiterungen oder die Visualisierung von Suchfacetten an. Die Qualität der Benutzerinteraktion wurde aufgrund folgender Subkriterien beurteilt: Präsentation Trefferliste, Nutzerführung und Performance.

Der Median von 0.42 sowie ein Maximum von 0.68 sind überraschend tief. Das ist insbesondere deswegen erstaunlich, da dieser Aspekt der Suchapplikation – da sichtbar und bedienbar – einfach, verständlich und testbar ist. Zudem lassen sich die meisten Anforderungen an die Benutzerinteraktion mit horizontalen Internet-Suchdiensten, die auf derselben Kollektion arbeiten, vergleichen. Zusammengefasst müsste eine Suchfunktion rasch eine gut rangierten Trefferliste auf die Anfrage erzeugen, diese visuell in einer erwartungskonformen Art präsentieren und zudem den User bei der Informationssuche aktiv unterstützen.

Die Bewertung der Subkriterien ist wie folgt (siehe Abbildung 5):

| Statistiken | Präsentation Trefferliste | Nutzerführung | Performance |
|-------------|------------------------------|---------------|-------------|
| Maximum | 0.91 | 0.63 | 1.00 |
| 3. Quartil | 0.76 | 0.32 | 1.00 |
| Median | 0.68 | 0.20 | 1.00 |
| 1. Quartil | 0.47 | 0.12 | 0.50 |
| Minimum | 0.12 | 0.06 | 0.00 |

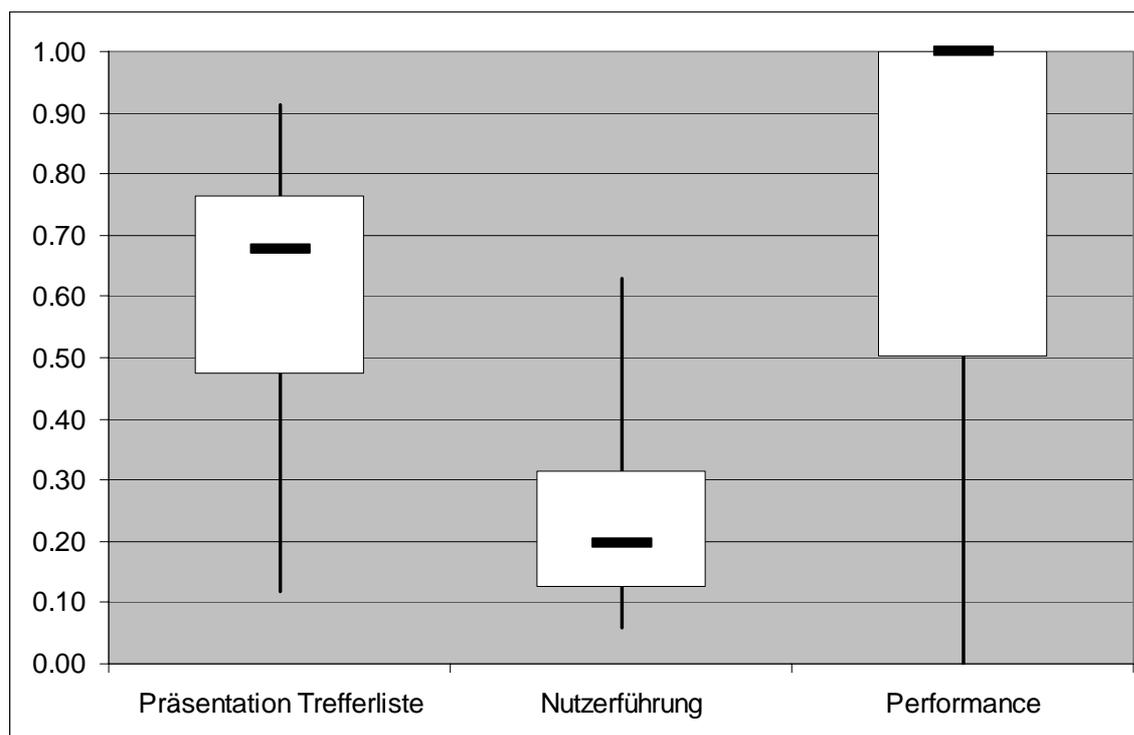


Abbildung 5: Hauptkriterium "Benutzerinteraktion": Boxplot der Subkriterien

Die Nützlichkeit der Trefferliste hängt stark vom Vorhandensein bestimmter Informationen und Funktionen, sowie von deren Darstellung ab. Diese Anforderungen wurden im Subkriterium **Präsentation der Trefferliste** geprüft. Die Nutzeranforderungen für eine öffentliche Site-Suche sind gut erforscht. Durch die intensive Nutzung von öffentlichen Suchdiensten haben Nutzer zudem Darstellungen sowie gewisse Funktionen erlernt. Davon abzuweichen ist bei ähnlichen gelagerten Anforderungen nicht zu empfehlen und löste bei den Testern Irritation aus. Im Bezug auf nicht funktionale Aspekte, der reinen Präsentation, zeigte der Test mit einem Median von 0.68 und einem unteren Quartil oberhalb 0.47 einen ziemlich guten Wert. Die Varianz über alle Tests ist dennoch sehr gross. Grund dafür sind anspruchsvoller umzusetzende Anforderungen wie beispielsweise die Qualität von Trefferzitate oder die Hervorhebung des Suchbegriffes im Treffer. Aber auch grundlegende Aspekte wie die Anzeige der URL oder des Pfades des Treffers, die Anzeige des Dokumentdatums oder der Grösse des Dokumentes wurden in einigen Fällen vernachlässigt. Bei den schlecht bewerteten Site-Suchen wurde der Suchfunktion offensichtlich eine geringe Bedeutung zugemessen oder eine Umsetzung ohne Nutzerfokus gewählt. Im Rahmen der Evaluation nicht bewertet, aber von den Testern als überflüssig angemerkt, wurde auf vielen Sites der geschätzte Relevanzwert in Prozent gezeigt. So wie beim Hauptkriterium Suchindex (vgl. Kapitel 4) war auch hier die Behandlung von binären Dokumentformaten, insbesondere von Adobe PDF, pro-

blembehaftet. Bei den Testkriterien der Qualität des Trefferzitates eines PDF-Dokumentes¹⁶, dem Vorhandensein von HTML-Versionen und der Hervorhebung der Suchworte im Trefferzitat ist über die Hälfte aller Suchanwendungen gescheitert.

Nutzerführung ist eine sehr wichtige Anforderung, um Nutzer bei der Suche zu unterstützen. Mit einem Median von 0.2 ist dieser Wert, mit Ausnahme der Aktualität des Suchindex (vgl. Kapitel 4), der schlechteste Wert eines Subkriterium der ganzen Evaluation. Dies, obschon es bei der Suche um die Nutzer geht und sich viele grundlegende Probleme der Informationssuche wie beispielsweise Subjektivität von Relevanz oder fehlendes Kontextwissen bei Suchbeginn exakt in diesem Feld bewegen. Bereits an der Anforderung, eine Suche mit der Eingabe-Taste auszulösen, scheiterten zwei der Websites. Bezüglich der Editierbarkeit des Suchbegriffs auf der Trefferliste schon acht. Anspruchsvollere Funktionen wie Korrekturvorschläge für Tippfehler oder eine aktive Unterstützung durch Erweiterung des Suchbegriffs sind bei fast allen Websites abwesend.

Die **Performance** ist in fast allen Fällen gut, wie der Median von 1 und die schiefe Verteilung zu Gunsten der Maximalbewertung zeigt. Im Rahmen des Tests zeigte sich klar, dass die Performance ein klarer Unzufriedenheitsfaktor ist. Ist die Antwortzeit gut, wird kaum gelobt. Ist dies aber ungenügend, so reagierten die Testpersonen stark irritiert bis hin zu Testabbrüchen. Bei den als langsam bewerteten Suchfunktionen liess sich ausserhalb des Testrasters gut beobachten, dass die Antwortzeit einen Zusammenhang mit der Grösse der Resultatsmenge hat.

Wie bereits argumentiert, sind die Anforderungen an eine Site-Suche nicht dieselben wie an einen horizontalen Internet-Suchdienst. In Bezug auf die Benutzerschnittstelle empfiehlt sich jedoch eine starke Anlehnung. Dies, da gewisse Interaktionsprinzipien durch die häufige Nutzung erlernt wurden und die visuelle Leitung dieselben Anforderungen zu erfüllen hat. Fast gänzlich abwesend sind hingegen moderne Arten der Informationserschliessung oder spezielle Lösungsansätze entlang der Kommunikationsziele der Suchfunktion.

¹⁶ In Tombros & Sanderson (1998) wird die Wichtigkeit von Trefferzitatzen aus Nutzersicht sowie Ansätze für die Erstellung diskutiert.

7. Suchergebnisse

Broder (2002) unterscheidet in seinem viel beachteten und zitierten Artikel zwischen Navigationsanfragen, Informationsanfragen und Transaktionsanfragen. Bei den **Navigationsanfragen** (navigational queries) geht es darum, eine Website oder eine Homepage zu finden, um von dort aus weiter navigieren zu können. Bei **Informationsanfragen** (informational queries) geht es um klassische Informationsrecherchen, bei der Informationen zu einem bestimmten Thema zusammengestellt werden müssen. Bei Transaktionsanfragen (transactional queries) soll der Einstiegspunkt für eine Online-Transaktion gefunden werden. Andrei Broder hat diesen Artikel geschrieben, bevor er von Altavista zu IBM wechselte und hatte Zugang zu den Altavista-Logdateien. Für unsere Zwecke haben wir die Taxonomie der Anfragen modernisiert, indem wir Transaktionsanfragen als eine spezielle Form von Navigationsanfragen betrachten und zusätzlich Faktenanfragen berücksichtigen. Bei **Faktenanfragen** soll ein Faktum gefunden werden, beispielsweise eine Adresse, ein Name oder eine Telefonnummer. Ein einziges vertrauenswürdiges Dokument genügt, welches das gesuchte Faktum enthält. Die anderen Dokumente interessieren dann nicht mehr, weil man das Informationsbedürfnis mit diesem einen Dokument befriedigen konnte.

Für unsere Bewertung haben wir die drei Anfragetypen gleich bewertet, da sie etwa gleich häufig vorkommen. Auch hier ergibt sich für die drei Subkriterien ein unterschiedliches Bild (siehe Abbildung 6):

| Statistiken | Navigationsanfragen | Informationsanfragen | Faktenanfragen |
|-------------|---------------------|----------------------|----------------|
| Maximum | 1.00 | 0.92 | 1.00 |
| 3. Quartil | 1.00 | 0.50 | 0.85 |
| Median | 0.75 | 0.25 | 0.70 |
| 1. Quartil | 0.69 | 0.06 | 0.40 |
| Minimum | 0.00 | 0.00 | 0.00 |

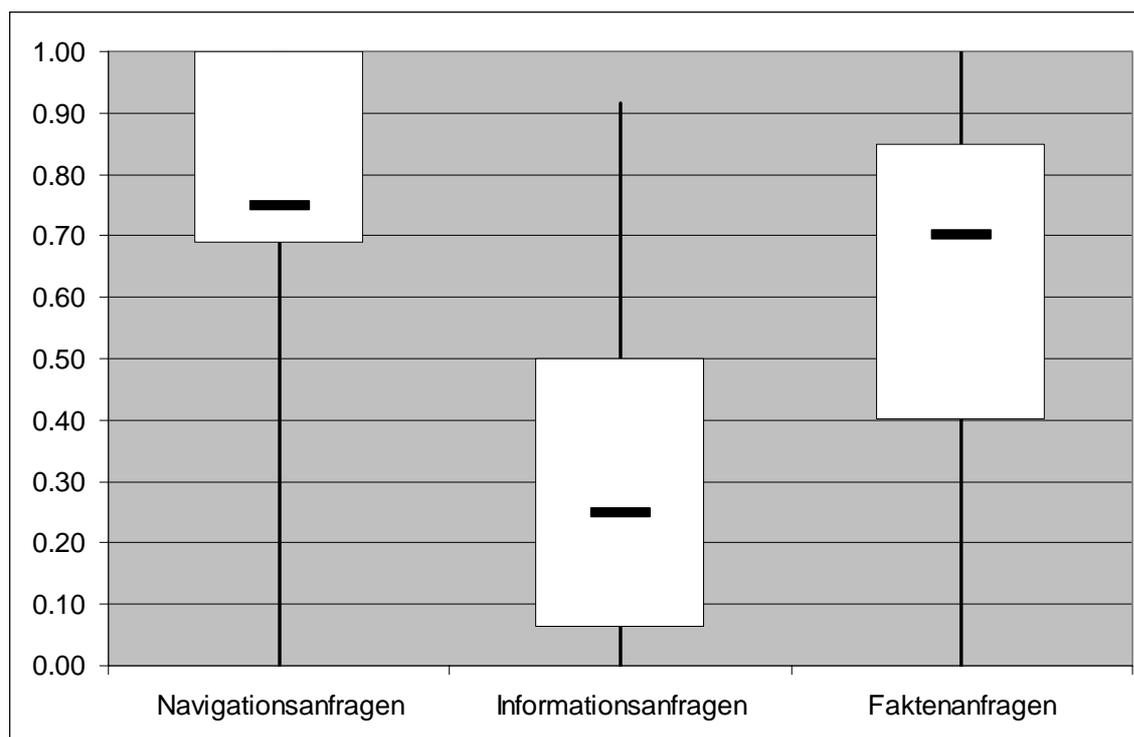


Abbildung 6: Hauptkriterium "Suchergebnisse": Boxplot der Subkriterien

Bei den Navigationsanfragen liefern die meisten Site-Suchen gute Ergebnisse, die restlichen jedoch schlechte bis sehr schlechte. Bei den Faktenanfragen liefert die Hälfte der Suchapplikationen gute Ergebnisse, ein Viertel der Applikationen kann als genügend bezeichnet werden. Der letzte Viertel ist klar ungenügend. In der klassischen Information Retrieval Disziplin, nämlich den Informationsanfragen, sieht es schlecht aus. Nur Ausreisser sind genügend, die meisten Suchapplikationen sind ungenügend bis schlecht.

Im Folgenden wird das Vergleichen von Anfragen und Dokumenten anhand einiger Beispiele erläutert:

- Die Anfrage nach offenen Stellen (Jobs, Stellen, etc.) ist eine typische Navigationsanfrage. Der Benutzer sucht den Einstiegspunkt, von dem aus er weiter im aktuellen Stellenangebot navigieren kann. Die Nutzungsanalysen von vielen Site-Suchen zeigen, dass die Frage nach offenen Stellen zu den am häufigsten gestellten Fragen gehört. Um so mehr erstaunt, dass einige Suchapplikationen kein befriedigendes Resultat liefern.
- Dass Faktenanfragen für eine Site-Suche nicht immer trivial zu beantworten sind, zeigt die Testanfrage nach der Anzahl der Mitarbeitenden. Ein Drittel der getesteten Applikationen konnten mit dieser Anfrage nichts anfangen. Der Hauptgrund liegt bei den verschiedenen Formulierungsmöglichkeiten. Gemäss Furnas et al. (1987)

besteht eine 80% Wahrscheinlichkeit, dass der Benutzer der Site-Suche und der Autor einer relevanten Website verschiedene Bezeichnungen benutzen.

- Die oben erwähnte Vielfalt der Möglichkeiten, ein Informationsbedürfnis als Anfrage zu formulieren und eine zu kommunizierende Information auf einer Web-Seite auszuformulieren, erschwert natürlich auch das Vergleichen von Dokumenten mit Informationsanfragen. Die obige Untersuchung wurde für englische Texte gemacht. Bei deutschen Texten mit vielfältigen Möglichkeiten, Komposita zu bilden und den vielen Flexionen ist noch eine höhere Mismatch-Quote zu erwarten. Hier sind moderne Retrieval-Techniken gefragt, um mit dieser Art von Unsicherheit fertig zu werden. So ermittelte die Studie von Braschler & Ripplinger (2004) eine bis über 60% bessere Retrievaleffektivität auf deutschen Texten, wenn Flexionen und Komposita normalisiert werden.

8. Empfehlungen

Die Studie hat versucht, die Suchfunktion von Unternehmens-Websites eingehend zu beleuchten, in dem vier Hauptkriterien identifiziert wurden, welche die Suchfunktion möglichst vollständig charakterisieren sollen. Diese Wahl hat sich als eine gute erwiesen, insofern es sich, wie teilweise auch erwartet, gezeigt hat, dass die Websites weit davon entfernt sind, konsistent in allen Hauptkriterien gut abzuschneiden. Diese Tatsache dient auch als weitere Illustration der Gefahren, Evaluationen auf einzelne Leistungsgrößen zu reduzieren – ein solcher „Leistungsscore“ mag in seiner Schlichtheit verführerisch sein, da er eine absolute Vergleichbarkeit der untersuchten Systeme suggeriert. In unseren Untersuchungen haben wir diesem Reiz bewusst widerstanden: es ist etliche Male überzeugend nachgewiesen worden, dass Evaluationstests von Retrievalsystemen deutlich zuverlässigere Resultate im Vergleich zwischen Systemen liefern als bei der Interpretation von absoluten Leistungsgrößen (Voorhees, 2002).

Das grosse Ziel der Studie wurde denn auch erreicht: der State-of-the-Art in Sachen Site-Suche für Unternehmen und Behörden in der Schweiz wurde ermittelt und analysiert. Das Fehlen von Systemen, die in allen Hauptkriterien brillieren, kann als Anreiz aufgefasst werden, eine reale existierende Leistung zu erreichen, die deutlich über den Suchfunktionen der heute besten Websites liegt. Das Potential wurde von den einzelnen Site-Suchen in der Untersuchung demonstriert.

Die Einzelanalyse der Hauptkriterien lässt einige Ansatzpunkte hervortreten. Darunter befinden sich insbesondere:

- Mehr Gewicht auf die Vollständigkeit und Aktualität des Suchindex. Die Suche ist nur so gut wie die Datenbasis. Viele Websites haben hier Verbesserungsbedarf. Eine mögliche Abhilfe besteht in einer aktiveren Einbindung der Suchfunktion in die Gesamtwebsite (aktive Aktualisierungen des Suchindex statt passives „Crawlen“, d.h. Sammeln der Dokumente durch die Suchapplikation)
- Bessere Pflege der Metadaten im Hinblick auf Einschränkungen bei komplexeren Suchen. Eine minimale Qualität (korrektes Datum, korrekte Titel) hilft, dass diesbezügliche Einschränkungen funktionieren. Trefferzitate müssen vollständig und prägnant sein. Wichtige Informationen müssen in Dateiformaten vorliegen, welche sauber verarbeitet werden können (die Tester bemängelten häufig Probleme im Umgang mit Adobe PDF-Dokumenten).
- Orientierung an bewährten Standards für die Benutzerinteraktion. Die Tester haben fast immer irritiert reagiert, wenn ihnen die Site-Suche intransparent

erschien. In Zukunft ist vermehrt zu untersuchen, mit welchen Vorstellungen an die Interaktion die Nutzer sich der Site-Suche nähern. Gegebenenfalls hilft die Orientierung an horizontalen Internet-Suchdiensten.

- Ausbau der Qualität des Vergleichs zwischen Anfragen und Dokumenten. Die korrekte Verarbeitung der natürlichen Sprache (Flexionen, Komposita) muss mehr Gewicht erhalten. Gegebenenfalls ist Mehrsprachigkeit zu beachten. Nutzer von Suchfunktionen auf Unternehmens-Websites sind auch ausbeuteorientiert.
- Trefferlisten sind nicht nur aus Dokumentensicht, sondern auch aus Nutzersicht aufzubereiten. Gewisse häufige Informationsbedürfnisse können vorweg genommen werden. Die Suchergebnisse für solche Anfragen können durch vorgängige Optimierung (andere Gewichtung/Stratifikation der Trefferliste, Entitätenerkennung u.a.) verbessert werden. Mit relativ kleinem Aufwand kann so bei einer grosse Anzahl von Einzelanfragen eine bessere Leistung erreicht werden.

Die Suche muss sich an der Unternehmens-Website orientieren. Dabei ist die Information geeignet zu strukturieren und anzubieten. Der Site-Suche gehört eine hohe Aufmerksamkeit gewidmet, die ihrer Rolle als Kommunikationsinstrument gerecht wird.

9. Kritische Würdigung der Studie

Im Nachhinein stellen sich die beiden Fragen „Wurden die richtigen Kriterien evaluiert?“ und zweitens, „Wurden diese Kriterien richtig evaluiert?“. Bei der ersten Frage muss berücksichtigt werden, dass weder ein Sitewettbewerb noch ein Produktwettbewerb mit Ranglisten beabsichtigt war (siehe Abschnitt 2.2). Dies rechtfertigt, weshalb man auf die Kriterien

- Kosten (einmalige und laufende),
- Sicherheit (Site-Suchen sind im Internet exponiert) und
- Organisation (anspruchsvolle/einfache Wartung)

verzichtete.

Die Wahl der verwendeten Kriterien (Suchindex, Vergleich von Anfragen und Dokumenten, Benutzerinteraktionen und Suchergebnisse) basiert auf den Erkenntnissen, welche in den grossen Information Retrieval Evaluationsforen TREC, CLEF und NTCIR gemacht wurden. Die mehr als zehnjährige Tradition dieser Evaluationsforen und die Tatsache, dass die internationale Forschungsgemeinschaft in diesen Foren gut vertreten ist, rechtfertigt die Auswahl. Unter den in dieser Studie verwendeten Kriterien gibt es keine offensichtliche Abhängigkeit, was die Auswahl sinnvoll erscheinen lässt.

Sehr gerne hätten die Verfasser der Studie die Retrieval-Effektivität umfassend ermittelt, wie dies in den oben erwähnten Evaluationsforen gemacht wird. Dabei werden in standardisierten Umgebungen mit speziellen Massen (Ausbeute, Präzision, P@10, etc.) ermittelt, ob eine Suchapplikation zu den Anfragen die relevanten Informationen *effektiv* findet. Da wir es bei den Site-Suchen nicht mit standardisierten Umgebungen zu tun haben, hätte man die Retrieval-Effektivität mit einem speziellen Experiment ermitteln müssen, wie dies im letzten Jahr an der ETH Zürich durchgeführt wurde. Bei 56 Website-Suchen hätte dies jedoch den Rahmen um Grössenordnungen gesprengt. Zukünftig ist eine derartige vertiefte Evaluation bei einzelnen Site-Suchen durchaus denkbar.

Die Frage, ob die verschiedenen Kriterien korrekt evaluiert wurden, muss differenziert beantwortet werden. Im Nachhinein konnten bei vereinzelt Tests Probleme identifiziert werden, welche zu nur teilweise korrekten Resultaten führten. Aufgrund der Erfahrung in den grossen Evaluationsforen wurde das Evaluationsraster jedoch so angelegt, dass diese Fehler keinen Einfluss haben auf die Kernaussagen. Die wichtigsten Mechanismen sind die Folgenden.

- Genügend grosse Anzahl von Tests, was mit 74 erfüllt ist.
- Die Testpersonen müssen die Tests unabhängig vom darunter liegenden Produkt evaluieren. Dies wurde erreicht, indem die Testpersonen keine Beziehung zu den Site-Suchprojekten hatten und die Zuteilung nach dem Zufallsprinzip erfolgte.
- Die Bewertungen erfolgen relativ zu den Bewertungen anderer Site-Suchen. Dies erfordert genügend viele zu testende Site-Suchen, was mit 56 erfüllt ist.

Zusammenfassend kann man sagen, dass die in der Studie gemachten Erkenntnisse fundiert sind. Die Kernaussagen werden nicht relativiert durch vereinzelte Unschönheiten, welche aber in zukünftigen Studien trotzdem eliminiert werden sollten.

10. Literaturverzeichnis

Alvarado, C., Teevan, J., Ackerman, M. and Karger, D. (2003). Surviving the Information Explosion: How People find their Electronic Information. AI Memo 2003-006. <http://haystack.lcs.mit.edu/papers/alvarado.aim03.pdf>, verifiziert am 28.11.2006, 20:52 Uhr MEZ

Braschler, M., Ripplinger, B. (2004). How Effective is Stemming and Decompounding for German Text Retrieval? *Information Retrieval*, Volume 7, Issue 3/4, 291-306, Kluwer Academic Publishers

Cleverdon, C. W. (1967). The Cranfield tests on index language devices. *Aslib Proceedings*, 19, 173-192. Reprinted in (Sparck Jones & Willett, 1997)

Cremers, Iris et al. (2006). It's Time To Update Site Search Functionality. *Best Practices*. Forrester Research, Inc.

Dumais, S., Cutrell, E., Cadiz, J.J., Jancke, G., Sarin, R., and Robbins, D. C. (2003). Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. *Proceedings of SIGIR '03*, 72-79, 2003. <http://research.microsoft.com/~sdumais/SISCore-SIGIR2003-Final.pdf>, verifiziert am 24.11.2006, 11:52 Uhr MEZ

Fairthorne, R.A. (1961). The mathematics of classification. *Towards Information Retrieval*, Butterworths, London, 1-10 (1961)

Furnas, Landauer, Gomez, Dumais: The Vocabulary Problem in Human-System Communication. *Commun. ACM* 30(11): 964-971 (1987)

Frakes, William Bruce & Baeza-Yates, Ricardo (1992). *Information retrieval: data structures and algorithms*. Englewood Cliffs, NJ : Prentice Hall

Harman, D. (1993). The Text Retrieval Conferences. In *Proc of the TIPSTER Text Program (Phase I)*, September 1993.

Harman, D. K. (1995). The TREC Conferences. In *Hypertext - Information Retrieval – Multimedia: Synergieeffekte Elektronischer Informationssysteme*, *Proceedings of HIM '95* (pp. 9-28), Universitätsverlag Konstanz

Haerst, M. A. (2006). Clustering versus Faceted Categories for Information Exploration. *Communications of the ACM* 49(4): 59-61 (2006)

Kando, Noriko (2004). „Evaluation of Information Access Technologies at the NTCIR Workshop. In Carol Peters, Julio Gonzalo, Martin Braschler, Michael Kluck (Eds.), Comparative Evaluation of Multilingual Information Access Systems. 4th Workshop of the Cross-Language Evaluation Forum, CLEF 2003, Lecture Notes in Computer Science, Vol. 3237 (pp. 29-43), Springer

Lyman, P. & Varian, H. R. (2003). How Much Information, 2003. <http://www.sims.berkeley.edu/how-much-info-2003>, verifiziert am 24.11.2006, 11:53 Uhr MEZ

Peters, C. & Braschler, M. (2001) European research letter: Cross-language system evaluation: The CLEF campaigns, Journal of the American Society for Information Science and Technology, Volume 52, Issue 12, 1067-1072, John Wiley & Sons

Van Rijsbergen, C. J. (1979). Information Retrieval, London: Butterworth

Schmidt-Mänz, N. & Bomhardt, C. (2005) Wie suchen Onliner im Internet? Absatzwirtschaft, Science Factory 2/2005.

Sparck Jones, Karen & Peter Willett, (1997). Readings in information retrieval. San Francisco, CA: Morgan Kaufmann

Tombros, A. & Sanderson, M. (1998). Advantages in query biased summaries in information retrieval. In Proceedings of SIGIR-98, Melbourne, Australia, Aug. 1998

WEMF-Report Plus (2006). MA NET WELLE 2/2006 (August 2006). Der User-Markt wächst weiter. http://www.wemf.ch/pdf/d/report/Report_plus_0602.pdf, verifiziert am 26.11.2006, 11:09 Uhr MEZ

Witten, Ian H.; Moffat, Alistair; Bell, Timothy C. (1999). Managing gigabytes : compressing and indexing documents and images. San Francisco, Calif. : Morgan Kaufmann

Voorhees, E. M. (2002). The Philosophy of Information Retrieval Evaluation. In C. Peters, M. Braschler, J. Gonzalo, M. Kluck (Eds.), Evaluation of Cross-Language Information Retrieval Systems, Second Workshop of the Cross-Language Evaluation Forum, CLEF 2001, Lecture Notes in Computer Science, Vol. 2406 (pp. 355-370)

11. Glossar

Eine **Website** bezeichnet den Netzauftritt einer Person oder Organisation. Eine Website besteht aus mehreren Webseiten und anderen Dokumenten (Dateien, Ressourcen), die durch eine einheitliche Navigation und Suchfunktion zusammengefasst und verknüpft sind.

Eine **Website-Suche** (oder kurz **Sitesuche**) bezeichnet die Suchfunktion, welche neben der Navigation als weitere Zugriffsmöglichkeit angeboten wird, um relevante Informationen in der Website zu finden. Eine Website-Suche besteht (1) aus Dokumenten, welche die Website anbietet, (2) aus einer Suchapplikation, d.h. Software sowie (3) der Konfiguration der Software.

Eine **Suchapplikation** ist eine Software, welche Dokumente erschliesst und zu Benutzeranfragen Suchresultate liefert mit Referenzen auf mögliche relevante Dokumente.

Eine **Retrieval-Applikation** ist eine Software, welche verschiedene Retrieval-Funktionen wie Suche, Kategorisieren, Informationen extrahieren, etc. unterstützt. Im Gegensatz zu einer Suchapplikation muss also eine Retrieval-Applikation nicht zwingend ein Suchfenster anbieten.

Ein **Internet-Suchdienst** besteht aus einer spezialisierten Suchapplikation, welche grosse Teile des Internet erschliesst und mit einem häufig erheblichen Administrationsaufwand als allgemein zugänglicher Internet-Dienst gratis angeboten wird. Seitdem das Web existiert, haben sich verschiedene Internetsuchdienste in ihrer Popularität abgelöst. Zu den bekannteren gehören Altavista, Excite, Yahoo und Google.

Bisher erschienene Schriften

Ergebnisse von Forschungsprojekten erscheinen jeweils in Form von Arbeitsberichten in Reihen.
Sonstige Publikationen erscheinen in Form von alleinstehenden Schriften.

Derzeit gibt es in den Churer Schriften zur Informationswissenschaft folgende Reihen:
Reihe Berufsmarktforschung

Churer Schriften zur Informationswissenschaft – Schrift 1
Herausgegeben von Josef Herget und Sonja Hierl
Reihe Berufsmarktforschung – Arbeitsbericht 1:
Josef Herget
Thomas Seeger
Zum Stand der Berufsmarktforschung in der Informationswissenschaft
in deutschsprachigen Ländern
Chur, 2004 (im Druck)
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 2
Herausgegeben von Josef Herget und Sonja Hierl
Reihe Berufsmarktforschung – Arbeitsbericht 2:
Josef Herget
Norbert Lang
Berufsmarktforschung in Archiv, Bibliothek, Dokumentation
und in der Informationswirtschaft: Methodisches Konzept
Chur, 2004 (im Druck)
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 3
Herausgegeben von Josef Herget und Sonja Hierl
Reihe Berufsmarktforschung – Arbeitsbericht 3:
Josef Herget
Norbert Lang
Gegenwärtige und zukünftige Arbeitsfelder für Informationsspezialisten
in privatwirtschaftlichen Unternehmen und öffentlich-rechtlichen Institutionen
Chur, 2004
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 4
Herausgegeben von Josef Herget und Sonja Hierl
Sonja Hierl
Die Eignung des Einsatzes von Topic Maps für e-Learning
Vorgehensmodell und Konzeption einer e-Learning-Einheit unter Verwendung von Topic Maps
Chur, 2005
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 5
Herausgegeben von Josef Herget und Sonja Hierl
Nina Braschler
Realisierungsmöglichkeiten einer Zertifizierungsstelle für digitale Zertifikate in der Schweiz
Chur, 2005
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 6
Herausgegeben von Josef Herget und Sonja Hierl
Reihe Berufsmarktforschung – Arbeitsbericht 4:
Ivo Macek
Urs Naegeli
Postgraduiertenausbildung in der Informationswissenschaft in der Schweiz:
Konzept – Evaluation – Perspektiven
Chur, 2005
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 7
Herausgegeben von Josef Herget und Sonja Hierl
Caroline Ruosch
Die Fraktale Bibliothek:
Diskussion und Umsetzung des Konzepts in der deutschsprachigen Schweiz.
Chur, 2005
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 8
Herausgegeben von Josef Herget und Sonja Hierl
Esther Bättig
Information Literacy an Hochschulen
Entwicklungen in den USA, in Deutschland und der Schweiz
Chur, 2005
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 9
Herausgegeben von Josef Herget und Sonja Hierl
Franziska Höfliger
Konzept zur Schaffung einer Integrationsbibliothek in der Pestalozzi-Bibliothek Zürich
Chur, 2005
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 10
Herausgegeben von Josef Herget und Sonja Hierl
Myriam Kamphues
Geoinformationen der Schweiz im Internet:
Beurteilung von Benutzeroberflächen und Abfrageoptionen für Endnutzer
Chur, 2006
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 11
Herausgegeben von Josef Herget und Sonja Hierl
Luigi Ciullo
Stand von Records Management in der chemisch-pharmazeutischen Branche
Chur, 2006
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 12
Herausgegeben von Josef Herget und Sonja Hierl
Martin Braschler, Josef Herget, Joachim Pfister, Peter Schäuble, Markus Steinbach, Jürg Stuker
Evaluation der Suchfunktion von Schweizer Unternehmens-Websites
Chur, 2006
ISSN 1660-945X

Über die Informationswissenschaft der HTW Chur

Die Informationswissenschaft ist in der Schweiz noch ein junger Lehr- und Forschungsbereich. International weist diese Disziplin aber vor allem im anglo-amerikanischen Bereich eine jahrzehntelange Tradition auf. Die klassischen Bezeichnungen dort sind Information Science, Library Science oder Information Studies. Die Grundfragestellung der Informationswissenschaft liegt in der Betrachtung der Rolle und des Umgangs mit Information in allen ihren Ausprägungen und Medien sowohl in Wirtschaft und Gesellschaft. Die Informationswissenschaft wird in Chur integriert betrachtet.

Diese Sicht umfasst die Teildisziplinen Bibliothekswissenschaft, Archivwissenschaft und Dokumentationswissenschaft. Auch neue Entwicklungen im Bereich Informationswirtschaft werden gezielt aufgegriffen und im Lehr- und Forschungsprogramm berücksichtigt.

Der Studiengang Information und Dokumentation wird seit 1998 als Vollzeitstudiengang in Chur angeboten und seit 2002 als berufsbegleitender Studiengang in Zürich. Künftig wird ein berufsbegleitender Masterstudiengang das Lehrangebot abrunden.

Der Arbeitsbereich Informationswissenschaft vereinigt Cluster von Forschungs-, Entwicklungs- und Dienstleistungspotentialen in unterschiedlichen Kompetenzzentren.

- Folgende Kompetenzzentren sind im Aufbau:
- Strategic Research
- Information Management & Competitive Intelligence
- Records Management
- Library Consulting
- Information Engineering Laboratory

Diese Kompetenzzentren sind im **Swiss Institute for Information Research** zusammengefasst.

IMPRESSUM

Verlag & Anschrift

Swiss Institute for Information Research (SII)

Hochschule für Technik und Wirtschaft HTW Chur
University of Applied Sciences
Ringstrasse / Pulvermühlestrasse 57
CH-7000 Chur
www.informationswissenschaft.ch
www.fh-htwchur.ch

ISSN 1660-945X

Studienleiter

Prof. Dr. Josef Herget
Telefon: +41 81 286 24 44
Email: Josef.Herget@fh-htwchur.ch

Sekretariat

Telefon : +41 81 286 24 24
Fax : +41 81 286 24 00
Email: clarita.decurtins@fh-htwchur.ch