

Ursula Hasler

Spurensuche: Nutzerverhalten auf der Site-Suche einer Hochschul-Website

Exemplarische Analyse der Suchanfragen auf der Website der ZHAW

Online-Kommunikation
Corporate Communications
**ZHAW Zürcher Hochschule
für Angewandte Wissenschaften**

Spurensuche: Nutzerverhalten auf der Site-Suche einer Hochschul-Website
Ursula Hasler

Zürcher Hochschule für Angewandte Wissenschaften

Online-Kommunikation

ISBN-10:

ISBN-13:

Alle Rechte vorbehalten

© Zürcher Hochschule für Angewandte
Wissenschaften, Winterthur

2009



Spurensuche: Nutzerverhalten auf der Site-Suche einer Hochschul-Website

Exemplarische Analyse der Suchanfragen auf der Website der ZHAW Zürcher Hochschule für Angewandte Wissenschaften

Ursula Hasler

Mit der Mitarbeit von Loig Roumois

Abstract

Die Suchbegriffe, die Nutzer bei einer Site-Suche (nicht Internet-Suche wie Google etc.) eintippen, sind unbeabsichtigte aber explizite Indizien für konkrete Informationsbedürfnisse der Nutzer und somit für die Website-Betreiber nutzbar zur Optimierung des Informationsangebots. Die vorliegende Studie analysiert als Pilotprojekt das Korpus der mit einem Nutzungsmonitor ausgewerteten Suchanfragen auf der Hochschulwebsite der ZHAW als Problem von Suchstrategie und Formulierstrategie. Dazu wird das Suchverhalten der Nutzer auf der Site-Suche mit dem Suchverhalten auf Internet-Suchen verglichen und die Suchanfragen bezüglich nachgefragter Themen und Anfragetypen als Grundlage für die Erfassung möglicher Informationsbedürfnisse analysiert. Der Untersuchungsgegenstand wird transdisziplinär betrachtet und Erkenntnisse aus der Informationswissenschaft (Information Retrieval, Nutzerverhalten), Kommunikationswissenschaft, Linguistik und Organisationskommunikation (Website als Kommunikations- und Marketinginstrument) verarbeitet. Die Auswertungen zeigen einen sehr hohen Anteil an Ein-Wort- Anfragen mit mehrheitlich Concrete Information Needs und zu mehrheitlich studienadministrativen Themen. Dies lässt auf einen hohen Anteil an Suchende mit Kenntnissen der Organisation schliessen. Die linguistische Analyse belegt auch, dass intuitiv und ohne Suchstrategie gesucht wird, gleichzeitig sind die Treffererwartungen von Suchenden mit Kenntnissen der Organisation aber hoch. Diese ersten Ergebnisse bestätigen, dass die retrieval-technologischen und die kommunikativen Anforderungen an eine Site-Suche eigentlich höher und komplexer sind als bei der Internet-Suche.

Inhalt

1. Ausgangslage	3
1.1 Informationssuche heute	3
1.2 Website einer Hochschule	4
1.3 Kommunikative Bedeutung der Suchfunktion auf der Website	4
2. Forschungsinteresse	6
2.1 Aktueller Stand der Forschung bzgl. Suchverhalten auf Internet-Suchmaschinen	8
2.2 Unterschiede zwischen Internet-Suche und Site-Suche	8
2.3 Formulieren von Informationsbedürfnissen: find what I mean not what I say	9
<i>Das Suchdilemma</i>	10
3. Vorgehen	11
3.1 Charakteristik des Korpus und Methodik	11
3.2. Untersuchungskriterien	11
4. Analyseergebnisse	12
4.1 Vergleich mit Internet-Suchen	12
<i>Anzahl Terme pro Anfrage</i>	12
<i>Verwendung von Operatoren</i>	13
<i>Nachfrage nach Personen</i>	14
4.2 Linguistische Analyse	15
<i>Anfragen mit Kurzformen</i>	15
<i>Orthographie und Rechtschreibung</i>	16
<i>Lexik: Getrennt und Zusammenschreibung</i>	17
4.3 Inhaltliche Analyse	19
<i>Themen</i>	19
<i>Anfragetypen</i>	22
<i>Null-Treffer-Anfragen</i>	26
5. Fazit und Ausblick	27
5.1 Unterschiede zu Internet-Suchen	27
5.2 Konkrete Informationsbedürfnisse und hohe Erwartungen	28
5.3 Kaum Suchstrategien	28
5.4 Optimierungspotenzial für die Kommunikation	29
5.5 Googlification	30
Literatur	32

Verzeichnis der Tabellen und Abbildungen

Tab. 1 Anzahl Terme pro Anfrage
Tab. 2 Suche mit Operatoren (+, -, *, „“)
Tab. 3 Details Operatoren
Tab. 4 Anfrageart (Suche nach Person oder Inhalt)
Tab. 5 Details bei Eigennamen
Tab. 6 Anfragen Inhalte
Tab. 7 Details Abkürzungen (interne Kenntnisse, relevant nur bei inhaltlichen Anfragen)
Tab. 8 Anzahl Schreibfehler (ohne Personenanfragen)
Tab. 9 Schreibweise von Komposita
Tab. 10 administrative und fachliche Anfragen
Tab. 11 Details zu administrativen Anfragen
Tab. 12 Details zu fachthemenatischen Anfragen
Tab. 13 Gesamtanteil studienthematische Anfragen
Tab. 14 Präzise und unspezifische Anfragen
Tab. 15 Anzahl transaktionsorientierter Anfragen
Tab. 16 Anzahl Terme pro Anfrage bei präzisen Anfragen
Tab. 17 Null-Treffer-Anfragen
Abb. 1 Auswertung in zwei Phasen
Abb. 2 Peak und long tail der Suchanfragen im untersuchten Sample
Abb. 3 Information Need und Request

1. Ausgangslage

1.1 Informationssuche heute

Die zentrale und von den Webnutzern am meisten verwendete Webtechnologie ist die Suchmaschine. War früher, noch vor kaum zehn Jahren (Google wird am 21.9.09 zehn Jahre alt), Suchen in elektronischen Datenbeständen und im Internet eine Aufgabe von ausgebildeten Informationsspezialisten oder Fachexperten mit Recherchekenntnissen, die sich mit Suchstrategien wie Gebrauch der Bool'schen Operatoren AND, OR, AND NOT auskannten, so ist die Situation heute eine komplett andere: alle suchen online, da immer mehr Informationen (immer mehr nur noch) online zugänglich sind, und alle suchen irgendwie, mit Halbwissen oder völlig intuitiv. Die Erfolgsgeschichte von Google beruht gerade darauf, dass diese Suchmaschine eine auf Suchlaien ausgelegte Usability aufweist. In den ersten Jahren unterstützte Google keine Bool'sche Operatoren; Mehrwortanfragen wurden defaultmässig mit AND verbunden (Notess, 2006), später führte Google gewisse Operatoren ein, um auch Suchende mit Kenntnissen von Suchstrategien zu bedienen (Notess, 2009). Je mehr die grosse Masse der Nutzer und ihre Bedürfnisse die Webtechnologie beeinflussen und je aktiver der Nutzer im Web wird (und sein darf - Paradigmenwechsel mit Web 2.0), desto mehr muss die Technologie mit dem intuitiven (Laien-)Nutzerverhalten interagieren. Google wurde von Informationsspezialisten vorgeworfen, mit seinen auf Suchlaien ausgerichteten Funktionalitäten die Suchenden „verdorben“ zu haben. Inzwischen ist die völlig intuitive Form des Suchens nach Informationen, die von Google ermöglicht wird, zur Standardanforderung für Internet-Suchmaschinen geworden. Dadurch vollzieht sich in der Nutzerforschung auch ein Paradigmenwechsel von der Präskription: wie sollte eine Anfrage gestellt werden, zur Deskription: wie stellen Nutzer effektiv ihre Anfragen. Bezüglich der Suchfunktionen auf Websites ergibt sich deshalb die Frage, ob Nutzer die Site-Suche genau gleich verwenden, d.h. ihr Suchverhalten von den Internet-Suchen auch auf die Site-Suchen übertragen. Das würde bedeuten, dass auch Site-Suchmaschinen sich an diesem Nutzer-Suchverhalten ausrichten müssen, obwohl die Anforderungen an die Suche auf Corporate Websites in einigen Punkten stark von den Anforderungen an Internet-Suchen abweichen (Braschler et al., 2006:7), was wir noch genauer untersuchen werden.

In der folgenden Analyse eines Korpus von Suchanfragen an eine Website wird deshalb untersucht, ob sich die Merkmale des Suchverhaltens, die bei Internet-Suchmaschinen festgestellt wurden, auch bei Site-Suchmaschinen nachweisen lassen. So ist zum Beispiel in Umfragen bei Nutzern von Internet-Suchmaschinen der Trend zu beobachten, dass die Nutzer für die Informationssuche immer weniger navigieren, nicht einmal auf der Ergebnisausgabeseite scrollen. Wird ein Treffer angeklickt und die gesuchte Information nicht gleich gefunden, wird auf der betreffenden Website selten durch Navigieren weitergesucht, sondern gleich zurück auf die Trefferliste der Suche geklickt (Schmidt-Mänz 2005 / Lewandowski 2007b). Die vorliegende Untersuchung ordnet sich also im Bereich Nutzerforschung ein, die das Verhalten der Nutzer im Web untersucht. Mögliche Untersuchungsmethoden in der Web-Nutzerforschung sind a) Beobachtung der Nutzer in Laborsituationen, b) Nutzerbefragungen und c) Analyse der Logfiles von Suchmaschinen. Da wir das Nutzerverhalten einerseits aus der Perspektive der Organisationskommunikation (welche Inhalte suchen die Besucher der Website) und andererseits mit einem linguistischen Blick (wie formulieren die Suchenden ihre Anfrage) untersuchen wollen, interes-

sieren uns in erster Linie die effektiven Spuren, die die Suchenden in der Suchmaschine hinterlassen, also die Analyse der Logfiles mit den Suchanfragen. Im Unterschied zu analogen Untersuchungen bei Internet-Suchmaschinen ist es einerseits einfacher, die Logfiles von Site-Suchmaschinen auszuwerten, andererseits schwieriger, an solche Daten heranzukommen, weil sie von den Besitzern der Websites als proprietär betrachtet werden. Die Funktion der Autorin als Leiterin der Online-Kommunikation der Gesamthochschule ermöglichte in diesem Fall den Zugriff auf diese Daten, die mit einer Zusatzsoftware zur Suchapplikation, einem Nutzungsmonitor, aufbereitet werden.

1.2 Website einer Hochschule

Die Charakteristik der Organisation, ihre Dienstleistungen und Produkte, ihre Zielgruppen, ihre Themen und ihre Organisationsform determinieren das Informationsangebot auf der Website. Es ist deshalb notwendig, über Hintergrundwissen zur Organisation zu verfügen, deren Korpus von Suchanfragen analysiert werden soll. Die ZHAW Zürcher Hochschule für Angewandte Wissenschaften ist 2007 aus einer Fusion von vier Hochschulen hervorgegangen, sie umfasst heute acht Departemente (Fakultäten) und bietet 29 Bachelor- und 11 Masterstudiengänge an. 1 800 Personen arbeiten an der ZHAW und über 8000 Studierende sind eingeschrieben. Dazu kommen noch über 2000 Teilnehmende von Weiterbildungsstudiengängen. Die ZHAW ist bezüglich Studienangebot und Anzahl Studierende die grösste Fachhochschule in der Schweiz. Die ZHAW-Angehörigen haben mit einem Login Zugriff auf ein geschlossenes Intranet mit internen Informationen, auch die Studierenden sind im Intranet mit Ausnahme von einigen Inhalten zugangsberechtigt. Die Studierenden einer Hochschule sind eine spezielle Nutzergruppe, weder Externe wie die Kunden noch Interne wie die Angestellten, ihre Stellung zur Organisation lässt sich am ehesten noch mit Mitgliedern eines Verbandes oder Vereines vergleichen.

Die Website der Hochschule wird zentral von einem Webgremium aus Vertretern aller Departemente und der Corporate Communications (Online-Kommunikation) der Hochschule betreut und weiterentwickelt. Das Informationsangebot ist hierarchisch gegliedert in einen sogenannten Dachbereich, der allgemeine Informationen über die gesamte Hochschule beinhaltet, und in acht vollständige Untersites für die Departemente, die zum Teil wiederum Untersites von Instituten beinhalten. Die gesamte Domain ZHAW umfasst ca. 25 000 Seiten, die von der Site-Suche indexiert werden. Es kann davon ausgegangen werden, dass der durchschnittliche Nutzer der Hochschul-Website und damit auch der Site-Suche eher jünger ist und über eine gute bis sehr gute Bildung verfügt, was sich eigentlich in der Analyse der Suchanfragen nachweisen lassen sollte. Aufgrund der „Produkte“ einer Hochschule kann auch vermutet werden, dass unter den Nutzern überdurchschnittlich viele Fachsuchende mit Rechercheerfahrung zu finden sind, was sich ebenfalls in den Suchanfragen zeigen müsste.

1.3 Kommunikative Bedeutung der Suchfunktion auf der Website

Je mehr Informationen vor allem oder nur noch online zugänglich sind, desto mehr Laien ohne jegliche Kenntnisse von Suchstrategien suchen Informationen mit Suchmaschinen

über online Kanäle (siehe oben). Parallel zur steigenden Bedeutung der Informationsvermittlung über online Kanäle und damit der zunehmenden Wichtigkeit ihrer Websites haben auch Website-Betreiber vor einigen Jahren angefangen, auf ihren Sites eine Suchfunktion anzubieten, damit die Inhalte besser auffindbar werden. Heute ist das Vorhandensein von meist proprietären Site-Suchen auf Websites von Unternehmen und Organisationen eine Selbstverständlichkeit. Noch vor wenigen Jahren aber gab es unter Usability-Experten eine grosse Kontroverse, ob eine Suchfunktion auf einer Website nötig und sinnvoll sei. Jared Spool schloss Ende der neunziger Jahre aus Beobachtungen in Usability-Tests, dass die Nutzer nur ein, zwei Versuche mit der Site-Suche machen und dann die Site verlassen, wenn sie das Gesuchte nicht gefunden haben. Spool vertritt noch heute die Ansicht, dass der Gebrauch des Suchfensters ein Hinweis auf schlechtes Webdesign und untaugliche Navigation sei (Qesenbery et al. 2008b:2). Jakob Nielsen hingegen vertrat immer die Ansicht, dass es auf jeder Webseite ein Suchfenster geben müsse, weil Usability-Untersuchungen zeigten, dass mehr als die Hälfte der Nutzer auf einer Website nicht Herumnavigieren wollten, sondern fokussiert seien und so schnell wie möglich eine bestimmte Information finden möchten. Die Suchfunktion hat für Nielsen deshalb zwei Funktionen, erstens dem Nutzer die Kontrolle über seinen Weg zur gewünschten Information zurückzugeben und zweitens als Rettungsanker zu dienen, wenn er sich in der Navigation völlig verhaspelt hat (Qesenbery et al. 2008b:1). Tatsache ist, dass das Vorhandensein eines Suchfensters auf einer Website heute eine Selbstverständlichkeit und seine Nutzung völlig normal ist.

Der oben erwähnte Trend, Informationen immer weniger durch Navigieren zu suchen und gleich die Suchfunktion zu verwenden, wird sich dadurch noch verstärken, dass heute auch auf allen Unternehmens-Websites Informationen via Suchfunktion gefunden werden können. Dies wird bestätigt durch eine Nutzerumfrage aus dem Jahre 2006 bei europäischen Online-Konsumenten, wie sie sich bei der Informationssuche auf einer unbekanntem Website verhalten: rund die Hälfte der Nutzer verwendet gleich die Suchfunktion, etwas mehr als ein Viertel sucht über Browsen (navigieren), die restlichen Nutzer suchen über Kontakt oder verlassen die Site, wenn sie das Gesuchte nicht sofort finden (Cremers 2006, in Braschler et al., 2006:5). Die Bedeutung der Site-Suche ist mittlerweile von der Organisationskommunikation und vom Marketing erkannt und Anbieter von Site-Suchen verwenden auch die entsprechende kommerzielle Argumentation: da hat man den Nutzer oder potenziellen Kunden mit viel Geld über AdWords, Bannerwerbung, Newsletters, Mailings etc. auf die Website gelockt und nun findet er unter Umständen nicht, was er sucht, weil keine oder eine schlechte Site-Suche ihn im Stich lässt. Also stellen sich für die Online-Kommunikation einer Organisation zwei grundlegende Fragen: erstens, herauszufinden was die Nutzer auf der Website überhaupt suchen, und zweitens, ob die von der Site-Suche auf eine bestimmte Suchanfrage ausgegebenen Treffer auch die richtigen, d.h. sinnvollen Ergebnisse sind. Eine solche qualitative Analyse bedingt, dass die Logfiles der Suchapplikation ausgewertet werden können und dass sie im besten Fall bereits aufbereitet vorliegen. Google analytics bietet die Funktion der Analyse der Website-Suche an; wer jedoch nicht unbedingt Google Zugriff auf die gesamten Traffic-Daten seiner Website geben möchte, muss die Logfiles manuell auswerten oder schafft sich eine Site-Suche an, die auch ein Nutzungsmonitoring anbietet.

2. Forschungsinteresse

Aus der skizzierten Ausgangslage stellen sich Fragen wie: Was charakterisiert sprachlich-kommunikativ eine Suchanfrage in einer Internet- oder Site-Suche? Unter welchen Bedingungen kann aus der Suchanfrage auf das Informationsbedürfnis geschlossen werden? Lassen sich aus den Suchanfragen bestimmte Nutzerverhaltensformen kategorisieren? Worin bestehen die typischen sprachlichen Phänomene von Suchanfragen, die das Information Retrieval der Suchmaschine verstehen muss?

Aus diesen Fragen leiten sich zwei Forschungsperspektiven für die vorliegende Pilotstudie ab. Einerseits eine praxisorientierte Auswertung der Suchanfragen mit dem Ziel, Hinweise auf Suchstrategien, Informationsbedürfnisse und Nutzerverhalten zu bekommen (Phase 1), um mittels der Resultate dann in einer zweiten Phase nach der Zusatzanalyse der Relevanz der gelieferten Treffer das Informationsangebot der Website zu optimieren (vgl. Abb. 1). Zweitens eine erkenntnisorientierte Analyse des Korpus der Suchbegriffe, um Indizien zu bekommen, welche Formulierstrategien Suchende unter der extremen durch das Suchfenster bedingten Reduktionsanforderung anwenden, um ihre Informationsbedürfnisse auszudrücken. Es geht dabei um sprachliche Ausdrucksformen der Nutzer, wenn sie eine Frage auf ein oder wenige Stichwörter reduzieren, d.h. ihr komplexes Informationsbedürfnis in wenige Wörter übersetzen oder transferieren müssen.

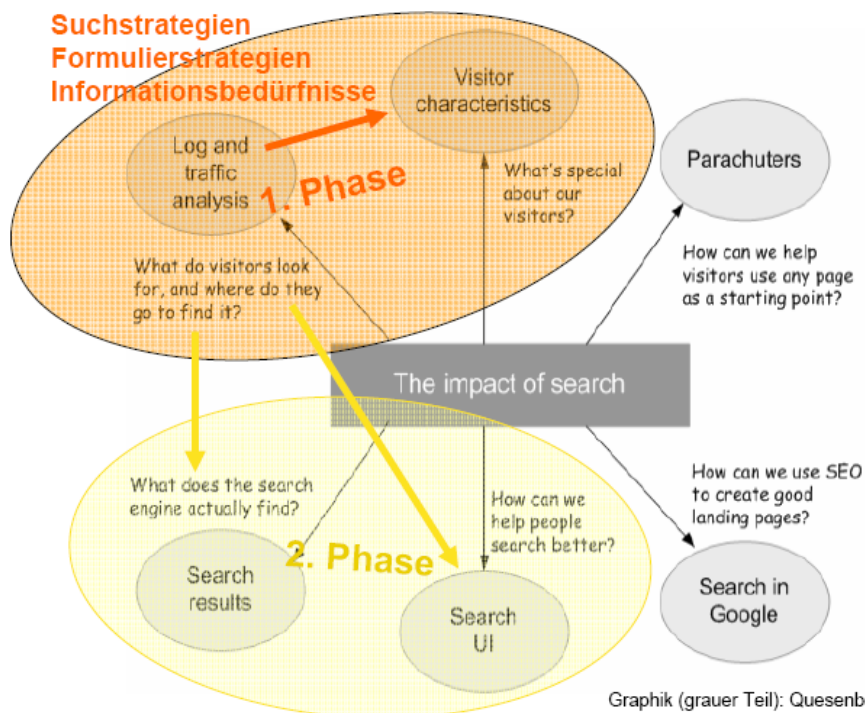


Abb. 1 Auswertung in zwei Phasen (Bearbeitung Graphik von Quesenbery)

Daraus ergeben sich aber teilweise unterschiedliche Interessen bezüglich des Untersuchungskorpus. Analysen von verschiedenen Samples von Suchanfragen haben gezeigt, dass es übereinstimmende Patterns bezüglich Suchwörter gibt: eine kleine Zahl von Suchwörtern, die innerhalb eines bestimmten Zeitraums sehr häufig eingegeben werden (sharp peak), eine sehr grosse Zahl von einmaligen Suchwörtern (long tail) und eine bestimmte Zahl von über lange Zeiträume immer wiederkehrenden Suchwörtern (persistent themes) (Quesenbery et al. 2008a:3, Höchstötter 2007:138). Dieses Muster zeigen auch die Suchanfragen dieser Studie in der untersuchten Woche (vgl. Abb. 2); die ersten zehn Ränge belegen folgende Suchanfragen: *stellen* wurde 102 mal gesucht, *stundenplan* 84 mal, *jobs* 73, *offene stellen* 70, *blackboard* 58, *militaer* 46, *haeberli* 36, *bibliothek* 32, *pruefungen* 28, *ferien* 28. Im „long tail“ sieht die Verteilung dann wie folgt aus: 51 Anfragen wurden 8 mal gestellt, 119 6 mal, 378 viermal und 1683 2 mal. Interessanterweise lässt sich dasselbe Pattern auch bei den Anfragen mit Null-Treffern nachweisen.

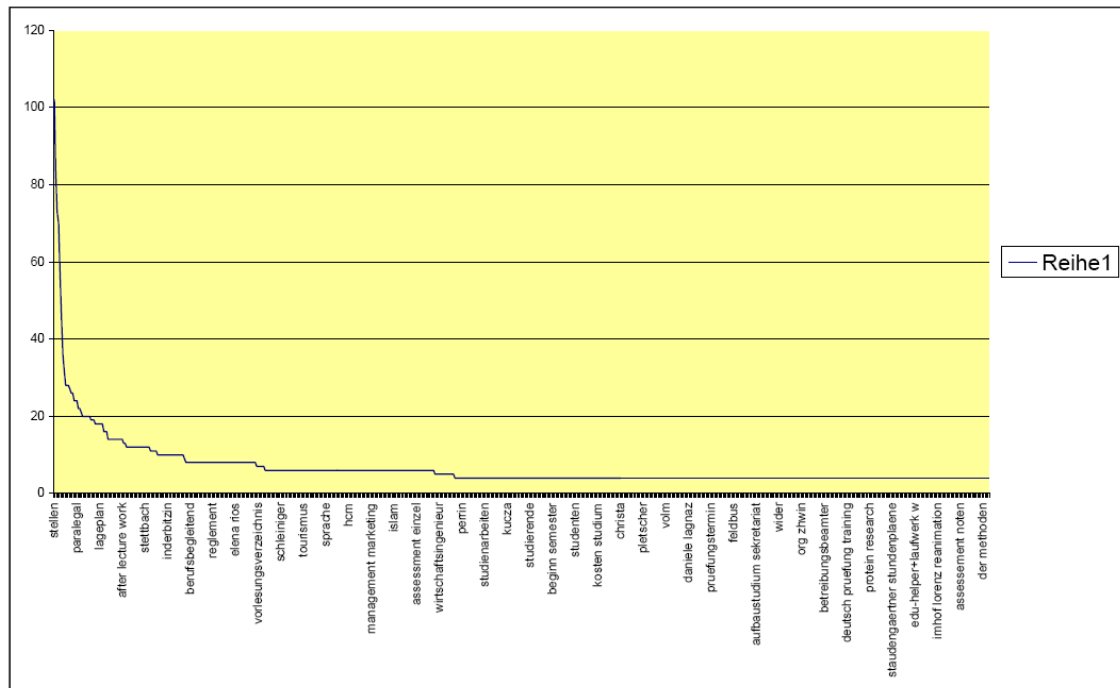


Abb. 2 Peak und long tail der Suchanfragen im untersuchten Sample

Aus der praxisorientierten Sicht der Informationswissenschaft resp. der Kommunikation reicht gemäss dem Pareto-Prinzip die Analyse der Suchanfragen im Peak (20%), um die wichtigsten Informationsbedürfnisse (80%) zu erkennen und allenfalls die Darstellung, Metadaten und Terminologie der entsprechenden Webseiten zu optimieren. Aus der erkenntnisorientierten Sicht der Linguistik hingegen ist für die Analyse der Formulierungsstrategien das Korpus sämtlicher Suchanfragen relevant, erst bei sprachlichen Auffälligkeiten wird die Häufung solcher Anfragen zusätzlich gewichtet. Für die nachfolgende Untersuchung wurde folglich das gesamte Korpus der Suchanfragen ausgewertet und erst in zweiter Linie auch die Häufigkeit noch berücksichtigt.

2.1 Aktuelle Erkenntnisse bzgl. Suchverhalten auf Internet-Suchmaschinen

In den vergangenen zehn Jahren wurden verschiedene Studien zur Nutzung der Internet-Suchmaschinen durchgeführt (Überblick in Schmidt-Mänz 2007a), und dabei je nach Forschungsinteresse auf unterschiedliche Aspekte fokussiert, beispielsweise Vergleich der Relevanz der Ergebnisse bei identischer Suchanfrage bei verschiedenen Suchmaschinen (u.a. Griesbaum 2004 / Véronis 2006 / Lewandowski 2009). Da für die vorliegende Analyse jedoch das Nutzerverhalten im Zentrum steht, sind insbesondere Untersuchungen der Suchstrategien der Nutzer wie Formulierung der Suchanfrage und Anfragetypen (u.a. Spink, Jansen 2004 / Lewandowski 2005, 2006 / Schmidt-Mänz 2007a,b / Fries 2007 / Höchstötter 2007) für einen Vergleich relevant. Nicht berücksichtigt wurden hier die jüngsten Studien zur möglichen Qualitätsmessung von Internet-Suchmaschinen, die die Anforderungen der Retrievalforschung und die Erkenntnisse aus Nutzerverhalten kombinieren (Stock, Stock 2008 / Lewandowski 2007a,b / Lewandowski, Höchstötter 2008). Ebenfalls wird für die vorliegende Pilot-Studie die Relation zwischen Suchanfrage und angeklickten Treffern noch nicht untersucht, dieser Aspekt soll jedoch in einer zweiten Phase nach Vorliegen der grundlegenden Erkenntnisse zu Suchanfragen auf Site-Suchen genauer analysiert werden, da er für die qualitative Untersuchung von Site-Suchmaschinen relevant ist und die Erkenntnisse für die Site-Betreiber einen direkten kommunikations- und marketingmässigen Nutzen aufweisen. Dabei sind die Usability und Qualität der Trefferausgabeseiten von Suchmaschinen zentral und müssen in die Untersuchung einbezogen werden (vgl. dazu Kaczmirek 2003, 2006 und Höchstötter, Lewandowski 2009).

2.2 Unterschiede zwischen Internet-Suche und Site-Suche

Es gibt bisher nur sehr wenige Studien zu Site-Suchen und Überlegungen zu den Unterschieden zwischen den beiden Suchen (Braschler et al. 2006 / Audrich 2006 / Quesenberry et al. 2008b). Internet-Suchen und Site-Suchen unterscheiden sich erstens grundsätzlich darin, dass die Datenmenge im Internet praktisch unbegrenzt ist, hingegen auf einer bestimmten Site klar durch den definierten Umfang der durchsuchten Domains begrenzt wird. Es wird deshalb von horizontaler (Internet-) resp. vertikaler (Website-) Suche gesprochen. Die begrenzte Datenmenge führt nun aber nicht automatisch zu besseren Suchresultaten, insbesondere bei unpräzisen Anfragen (lexikalische Vagheit) ist aufgrund der riesigen Informationsmenge die Trefferwahrscheinlichkeit in der Internet-Suche viel höher als auf einer Site-Suche und entsprechend auch die Nutzerzufriedenheit (Lewandowski 2007a), da immer etwas gefunden wird.

Zweitens weist jede Website eine spezifische Terminologie auf, deren Kenntnis oder Unkenntnis für den Suchenden entscheidend die Brauchbarkeit der Suchresultate beeinflusst (siehe oben). Während die Kenntnis resp. Unkenntnis einer bestimmten Terminologie bei Internet-Suchen aufgrund der grossen und auch redundanten Datenmenge weniger ins Gewicht fällt, ist die Frage „Wie wird xx auf dieser Website benannt?“ für den Suchenden bei einer Site-Suche zentral. Hier stellt sich die Frage, ob sich die terminologische Unsicherheit in der Formulierung der Suchanfrage nachweisen lässt.

Drittens kann bei Site-Suchen im Gegensatz zu Internet-Suchen von gewissen Hypothesen über die Nutzergruppen ausgegangen werden. Die Wahrscheinlichkeit, dass Personen, denen die Organisation oder ihre Geschäftsfelder bekannt sind, Suchanfragen formulieren, ist relativ gross, wobei der Anteil dieser Nutzer an der Gesamtmenge von Nutzern, die Suchanfragen stellen, natürlich je nach Art von Organisation und Typ von Website schwankt. Nutzer mit Kenntnis der Organisation haben präzise Informationsbedürfnisse, was jedoch nicht unbedingt bedeutet, dass sie auch die organisationspezifische Terminologie gut kennen.

Viertens ist der Einsatz von dialogischen Elementen zur Einschränkung der Anfrage oder andere interaktive Möglichkeiten zur Präzisierung des Informationsbedürfnisses bei Site-Suchen einfacher als bei Internet-Suchen, da die Kommunikations- oder Marketingabteilungen über Kenntnisse der Nutzergruppen verfügen. Die aktuellen Trends bei Suchmaschinen zur Verbesserung der Trefferqualität betreffen alle die Interaktion mit dem Nutzer (u.a. Web 2.0 Features, Google suggest, weiterführende thematische Linklisten, Disambiguation von Begriffen bei Wikipedia). Für die Entwicklung von dialogischen Elementen kann gezielt die Kompetenz der Organisationskommunikation (Online-Kommunikation) und ev. des Marketings zum Einsatz kommen.

2.3 Formulieren von Informationsbedürfnissen: find what I mean not what I write

Im Information Retrieval wird unterschieden zwischen *Real Information Need* (subjektives Informationsbedürfnis), *Perceived Information Need* (die Wahrnehmung des Informationsbedürfnisses durch den Nutzer), *Request* (Suchanfrage, die Formulierung des wahrgenommenen Informationsbedürfnisses durch den Nutzer in natürlicher Sprache) und *Query* (formalisierte Repräsentation der Suchanfrage in der Systemsprache, definiert den Informationsbedarf).

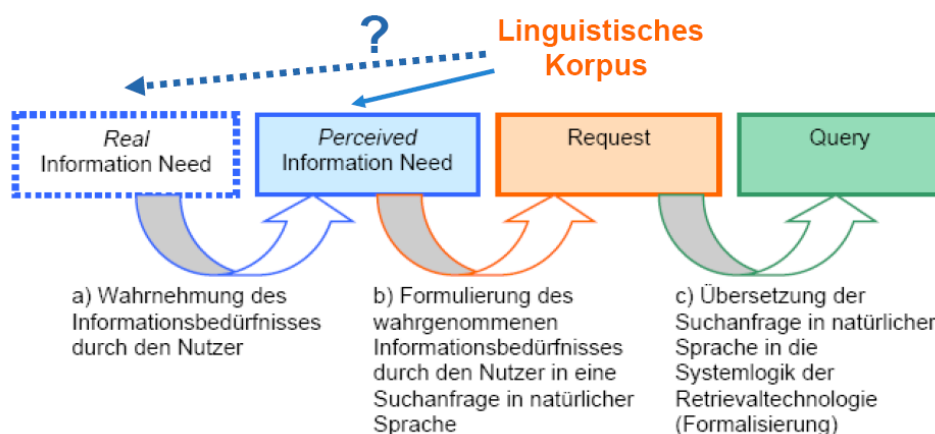


Abb. 3 Information Need und Request

Die Übergänge zwischen dem Informationsbedürfnis und seiner Wahrnehmung (a), der Wahrnehmung und der Formulierung (b) und der Formulierung und der Formalisierung (c) sind problematisch, da es keinen äquivalenten Transfer gibt, sondern nur Interpretati-

on. Aus Sicht der Nutzerforschung interessieren die Transformationen (a) und (b); (c) ist eine Frage der Retrievalforschung. Für den Suchenden stellt sich bei der Wahrnehmung des subjektiven Informationsbedürfnisses (a) das Hauptproblem, dass er etwas suchen muss, dass er unter Umständen nicht kennt oder wovon er nur vage Vorstellungen hat. Er schätzt sein Informationsproblem eventuell falsch ein und hat folglich auch unklare Vorstellungen über das Ergebnis, das sein Informationsbedürfnis befriedigt. Die sichtbare Spur seines Verständnisses des Informationsproblems ist dann die Formulierung seines Informationsbedürfnisses (b): Präzision oder Vagheit, Schlüsselwörter oder Phrasen, mit Kenntnis der Suchmaschinen-Formalisierung (Einsatz von Operatoren) oder „intuitiv“, etc. Das Hauptproblem beim Formulieren einer Suchanfrage besteht jedoch darin, dass der Suchende dem Begriff keinen Kontext mitgeben kann, der seinen Begriff thematisch in einen Kontext einordnet, mit dem er seine Intention ausdrücken kann. Suchexperten können bis zu einem gewissen Grad die Booleschen Operatoren dafür nutzen, um die in Frage kommenden Informationsquellen genauer zu bestimmen, unklar ist aber, welche Strategien Suchlaien anwenden. Folglich stellt sich hier die weitere Frage, inwiefern sich das Sucher-Dilemma in der linguistischen Analyse der Suchanfragen nachweisen lässt und mit welchen sprachlichen Mitteln (Kotext, Kontext, Anweisungen?) Suchlaien versuchen, ihre Intention auszudrücken. Ausgehend von diesen Antworten kann dann untersucht werden, inwiefern die Suchenden mit interaktiven Elementen unterstützt werden könnten, um die Formulierung zu präzisieren und ein allenfalls unklares Informationsproblem besser zu erfassen.

Das Suchdilemma

Eine der Hauptschwierigkeiten beim Formulieren des Informationsbedürfnisses als Suchanfrage bei einer Suchmaschine besteht darin, dass gar keine Frage formuliert werden kann. Der Suchende muss eine einfache Frage wie z.B. „Wie komme ich mit dem öffentlichen Verkehrsmittel zu diesem Institut?“ (Perceived Information Need) sprachlich in eine andere Logik übersetzen. Er kann der Suchmaschine nicht seine Frage eingeben, weil er nicht diese Frage selber, sondern die Antwort in den Informationsquellen sucht. Bei Internet-Suchen hingegen macht die Eingabe einer Frage manchmal Sinn, nämlich dann, wenn als Treffer ein Forum gesucht wird, wo diese Frage gestellt wurde. Das Hauptproblem für den Suchenden besteht darin, dass er wissen muss, was seine Frage beantworten könnte, und dass er danach suchen muss. Er muss also seine Frage transferieren in „Was (welche Infoquellen auf dieser Website) könnte meine Frage beantworten?“ Dies ist alles andere als trivial und verlangt vom Suchenden Zusatzwissen fachlicher, organisationsspezifischer oder suchspezifischer Art. Er muss zum Beispiel wissen, dass die Information, die seine Frage eventuell beantwortet, auf Websites in der Regel auf Seiten zu finden ist, die mit *Lageplan*, *Situationsplan*, *Anfahrt*, ev. auch *Kontakt* etc. betitelt sind und er folglich am besten einen dieser Begriffe ins Suchfeld eingibt, aber eher nicht Begriffe wie *öffentliche Verkehrsmittel* oder *Bus* oder *Bahn*. Der erwähnte Trend zur Interaktion mit dem Nutzer nimmt genau diese Problematik auf, indem versucht wird, nach Analyse einer Suchanfrage dem Suchenden Ergänzungsfragen zu stellen und so einen Kotext herzustellen, damit der richtige Kontext ausgewählt wird.

3. Vorgehen

3.1 Charakteristik des Korpus und Methodik

Die Site-Suche auf der Website der ZHAW Zürcher Hochschule für Angewandte Wissenschaften ist eine proprietäre Software, Relevancy Retrieval von Eurospider Information Technology AG in Zürich. Die Logfiles werden mit einem Monitoringtool, dem Relevancy Usage Monitor, bezüglich Anfrageart und -häufigkeit, pro Anfrage ausgegebenen Ergebnislisten (Links auf Quellen), angeklickten Treffern, zeitlicher Verteilung und Null-Treffern statistisch ausgewertet. Der Nutzungsmonitor bietet zusätzlich die Möglichkeit, häufige und ähnliche Anfragen manuell zu clustern und dem Nutzer nach Eingabe einer solchen Suchanfrage eine optimale Seite als Ergebnis vorzuschlagen.

Basis für die Pilotstudie ist die Gesamtheit aller Suchanfragen, die 2009 während einer Woche in die Suchfelder der Domain www.zhaw.ch und alle Unterdomains eingegeben wurden. Das Korpus ist folglich ein Abbild für schriftsprachliches Formulieren und belegt den Sprachgebrauch einer mit grosser Wahrscheinlichkeit eher jüngeren Nutzergruppe mit höherer Bildung (Zielgruppen einer Hochschule: Studierende, Interessenten, wissenschaftliche Bezugsgruppen etc.). Inhaltlich bildet die Charakteristik der Hochschule (Struktur, Art) den Kontext, thematisch eingeschränkt auf das Tätigkeitsfeld der Organisation (Aktivitäten, Dienstleistungen, Produkte, Fachbereiche, hier: acht Departemente mit mehreren Disziplinen).

3.2. Untersuchungskriterien

Entsprechend den unterschiedlichen Untersuchungsperspektiven wurden drei Gruppen von Kriterien gebildet:

1. Kategorien für den Vergleich von Nutzerverhalten auf Internet-Suchen und auf Site-Suchen:

1. Anzahl Terme pro Anfrage
2. Verwendung von Operatoren
3. Anfrageart (Suche nach Person oder Inhalt)

2. Kategorien für die linguistische Analyse der Formulierung der Anfrage

4. Kurzformen
5. Orthographie und Rechtschreibung
6. Lexik: Getrennt- und Zusammenschreibung

3. Kategorien für die inhaltliche Analyse der Anfragen

7. Themen
8. Anfragetypen

4. Analyseergebnisse

Für die Pilotuntersuchung wurden sämtliche Suchanfragen ausgewertet, die in der Woche 24, vom 8.-14.6.09, an die Site-Suche von www.zhaw.ch gestellt wurden. Es waren in dieser Woche gesamthaft 8044 Suchanfragen, davon 2469 unterschiedliche Anfragen mit insgesamt 11 843 verschiedenen Termen (keine Stoppwörter).

4.1 Vergleich mit Internet-Suchen

Anzahl Terme pro Anfrage

Bei den zahlreichen Nutzerstudien zu den Internet-Suchen ist die Anzahl Wörter pro Suchanfrage immer ein wichtiges Untersuchungskriterium. Im Information Retrieval wird ein Wort als Zeichenfolge zwischen zwei Leerzeichen definiert, was in der Linguistik als graphematisches Wort bezeichnet wird (Fuhrhop 2008). Was retrievaltechnisch als Wort verstanden wird und wie die Suchmaschine damit umgeht, ist für die Bedeutungser-schliessung sehr wichtig, insbesondere bei den im Deutschen sehr häufig vorkommen-den Komposita. Die Verwendung von Komposita in den Suchanfragen wird deshalb bei 4.2 untersucht.

Tab. 1 Anzahl Terme pro Anfrage

Terme pro Anfrage	Anfragen	% - Anteil		% - Anteil ohne Personen-suche	Vergl. Internet-Suche Ø % D	Vergl. Internet-Suche Ø % E ¹	Anzahl Terme absolut
1	5289	65,78	66	70 (4154)	50	15	5289
2	2052	25,49	25	20 (1177)	?	32	4104
3	485	6,02	6	7 (430)	?	27	1455
4	136	1,70	2	2 (119)	?	15	544
5 und mehr	53	1,01	1	1 (82)	?	11	451
Spezialfälle	29						
Total	8044	100	100	100 (5962)	Ø 1,7 pro Anfrage ²	Ø 2,6 pro Anfrage ³	11843 ./ 8044 = Ø 1,47 pro Anfrage

Auffällig ist der hohe Prozentsatz an Ein-Wort-Anfragen (66%), der sogar 70% ausmacht, wenn er um den grossen Anteil an Suchanfragen nach Personen (siehe Tab. 2) bereinigt wird, da Suchanfragen nach Personen häufig aus zwei Wörtern bestehen. Auswertungen der Anzahl Terme pro Anfrage bei Internet-Suchen zeigen markante Unterschiede zwischen deutschen (Ø 1,7 Terme) und englischen (Ø 2,6 Terme) Internet-Suchen, die aber erklärbar sind. Da im Deutschen und im Englischen andere Regeln für Getrennt- resp. Zusammenschreibung gelten, hat dies statistische Auswirkungen auf die Anzahl Wörter pro Anfrage. Untersuchungen von englischen und deutschen (oder anderssprachigen)

¹ OneStat.com 2007 http://www.onestat.com/html/aboutus_pressbox56-word-phrases-in-search-engines.html

² Lewandowski, Höchstötter 2008

³ Spink, Jansen 2004

Suchanfragen können deshalb bezüglich der Anzahl Wörter nicht miteinander verglichen werden. Dies bestätigt auch die Auszählung auf der Site-Suche hier: Bei den Anfragen mit 4 und 5 Wörtern ist die Mehrzahl der Anfragen englisch und betrifft oft Kursbezeichnungen wie „master in banking and finance“ oder „mas supply chain manager“. Der Vergleich der Anzahl Wörter pro Suchanfrage in deutschen Internet-Suchen (\bar{x} 1,7) und auf der Site-Suche (\bar{x} 1,47) bestätigt einerseits, dass die Sprache ausschlaggebend ist, und wirft gleichzeitig die Frage auf, ob der hohe Anteil an Ein-Wort-Anfragen, generell höher als bei Internet-Suchen, typisch für Suchanfragen auf Websites sein könnte, als Ausdruck dafür, dass der Nutzer praktisch sicher ist, dass das Gesuchte vorhanden sein muss.

Verwendung von Operatoren

Die Gesamtmenge aller Suchanfragen wurde auch bezüglich Verwendung von Operatoren ausgewertet:

Tab. 2 Suche mit Operatoren (+, -, *, *, „“)

Operatoren	Anfragen	%-Anteil gesamt		Vgl. Internet-Suche D⁴	Vgl. Internet-Suche E⁵	Bemerkungen
mit	206	2,56	3	< 3	10	Korrekte resp. sinnvolle Verwendung nur 1,82 %
ohne	7838	97,43	97			
Total	8044	100	100			

Die allgemeine Beobachtung, dass Kenntnisse über komplexere Suchstrategien wie erweiterte resp. eingeschränkte Suche bei einem breiten Publikum nicht (mehr) vorhanden sind, bestätigen diese Ergebnisse eindeutig. Nur 2,56% der Suchenden verwendet überhaupt Operatoren und ein Drittel von ihnen, wie die Detailanalyse Tab. 7 zeigt, auch nicht unbedingt mit Sachkenntnis. Diese Ergebnisse decken sich mit den Auswertungen der Logfiles von deutschen Internet-Suchen, aber nicht mit der Selbsteinschätzung der Nutzer selber, siehe Anmerkung (4). Da bei Internet-Suchen die Hälfte der Suchanfragen mit Operatoren logische Fehler enthielten, wurden die hier verwendeten Operatoren noch genauer untersucht:

⁴ Schmidt-Mänz, Koch 2006, Auswertung der Logfiles von Fireball, Lycos und Metaspinner. Diese Resultate divergieren extrem von den Aussagen der Nutzer über ihr Suchverhalten: Bei Befragungen gaben 52,7% (!) an, Operatoren zu benutzen, Schmidt-Mänz, Bomhardt 2005.

⁵ Spink, Jansen 2004, davon sind die Hälfte der Anfragen mit Operatoren fehlerhaft

Tab. 3 Details Operatoren

Art	Anfragen	alle %	unkorr.	korr.	korr %	Bedeutung der Operatoren
+wort (Pi)	133	64,56	33	100	68	Erzwingen: Alle Begriffe mit + (ohne Leerschlag vorangestellt!) müssen im Suchergebnis vorkommen (Irrelevant bei Google, da dies immer so ist, nur relevant bei Stoppwörtern).
-wort (Mi)	8	3,88	2	6	4	Ausschliessen: Alle Begriffe mit – (ohne Leerschlag vorangestellt!) dürfen im Suchergebnis nicht vorkommen.
wort(stamm)* (St)	11	5,33	8	3	2	Trunkieren: das Sternchen* markiert einen Platzhalter für eine Endung, alle Begriffe, die diesen Wortstamm enthalten, müssen im Suchergebnis vorkommen (wird von Google nicht unterstützt).
„phrase“ (ASZ)	54	26,21	16	38	26	Phrasensuche: alle Begriffe innerhalb von Anführungs-/Schlusszeichen müssen als Wortgruppe im Suchergebnis vorkommen.
Total	206	100	59	147		Rund 30% (59 von 206) sind fehlerhaft

Die Fehlerquote liegt mit rund 30% etwas tiefer als der bei Internet-Suchen ermittelte Anteil (50%), dies kann jedoch aufgrund des kleinen Untersuchungskorpus der Site-Suche zufallsbedingt sein. Signifikant ist, dass die korrekte Verwendung von Operatoren bei nur 1-2% der Anfragen minimal und somit vernachlässigbar ist.

Nachfrage nach Personen

Bei einer Site-Suche sind häufige Suchanfragen nach Personen zu erwarten, die in der Organisation arbeiten, deshalb wurden diese Anfragen ausgewertet:

Tab. 4 Anfrageart (Suche nach Person oder Inhalt)

Art	Anfragen	% - Anteil gesamt		Bemerkungen
Eigennamen (N)	2072	25.75	26	
Inhalte (I, A, X, F)	5972	74.24	74	Inkl. Abkürzungen (A), Unklares (X), Schreibfehler (F)
Total	8044	100	100	

Die Auswertung bestätigt die Vermutung und zeigt sogar einen überraschend hohen Anteil: nämlich ein Viertel aller Anfragen suchen nach Personen. Dieser grosse Prozentsatz könnte für Hochschulwebsites charakteristisch sein, da der grösste Teil der Hochschulangehörigen im Unterschied zu einem Wirtschaftsunternehmen stark extern orientiert arbeitet und viele Netzwerke ausserhalb der Hochschule (Scientific Community, Forschungszusammenarbeit mit Unternehmen, Dienstleistungen für die Öffentlichkeit etc.) pflegt. Hier zeigt sich ein weiterer Unterschied zwischen Site-Suche und Internet-Suche:

da die Site-Suche in der begrenzten Informationsmenge einer Organisation sucht, wird sie häufig benutzt, um die Kontaktdaten von Personen in der Organisation zu finden.

Tab. 5 Details bei Eigennamen

Wörter	Anfragen	%-Anteil gesamt		Bemerkungen
1	1135	54.67	54	Davon 55 Anfragen nur mit Vornamen!
2	879	42.34	42	16 Anfragen bei den 2-Wort-Anfragen haben die Form <i>Frau X</i>
3	53	2,55	3	Bei den 3- und Mehr-Wortanfragen beinhalten bei 22 Anfragen die Wörter Titelangaben <i>Dr. Prof.</i> etc.
4 und mehr	17	0,81	1	Bei den 3- und Mehr-Wortanfragen beinhalten 27 Anfragen einen Namen und eine Zusatzinformation wie <i>Hans Muster Diplomarbeit</i>
Total	2076	100	100	

Eher unerwartet gibt es auch bei den Personenanfragen mehr Ein- (54%) als Zwei-Wort-Anfragen (42%). Die Detailanalyse der Personenanfragen zeigt, dass die Sitesuche auch von Internen für die Personensuche benutzt wird, obwohl im Intranet eine präzisere Personensuche möglich wäre. Die Studierenden repräsentieren bei den Hochschulen eine spezielle Nutzergruppe bezüglich Zugang zu internen Informationen und internen Kenntnissen der Organisation. Bei der ZHAW verfügen sie über Zugangsberechtigungen zum Intranet der Hochschule (mit wenigen Einschränkungen), nutzen es jedoch wenig, vermutlich wegen des erforderlichen Logins. Der spezielle Status von Studierenden zwischen Externen und Internen oder auch als Halbinterne zeigt sich in den Anfragen, die „Halbwissen“ oder Bequemlichkeit ausdrücken. Es kann vermutet werden, dass die Suche nach Personen nur mit Vornamen oder *Frau X* oder auch die Verwendung von Namen mit Titeln von Studierenden stammen.

Die ZHAW führt seit mehreren Jahren eine sogenannte Porträt Datenbank, in der von allen ZHAW-Angehörigen (Staff) bei Eintritt die minimalen publizierbaren Kontaktdaten angelegt werden. Das Porträt kann nach Belieben von der Person mit Lebenslaufinformationen und wissenschaftlichen Leistungen (Publikationen, Projekte, Mitgliedschaften etc.) erweitert werden. Die Site-Suche Relevancy wurde so programmiert, dass bei der Suchanfrage nach einer Person als erster Treffer deren Porträt angezeigt wird. Die hohe Anzahl an Suchanfragen nach Personen zeigt, dass durch diese Massnahme der Kommunikation (Priorisierung der Porträt Datenbank in der Site-Suche) mindestens ein Viertel der Nutzerbedürfnisse optimal befriedigt werden können.

4.2 Linguistische Analyse

Anfragen mit Kurzformen

Die grosse Menge der inhaltlichen Anfragen wurde noch differenziert und Kurzformen sowie Unklares (Eingabefehler, abgebrochene Anfragen, Testanfragen etc.) ausgesondert:

Tab. 6 Anfragen Inhalte

Art	Anfragen	% - Anteil		Bemerkungen
Kurzformen (A)	694	11,62	12	Akronyme und interner Jargon
Unklares (X)	104	1,74	2	Einkopierte urls, Text im Suchfenster etc.
Übrige lexikalische Begriffe (I, F)	5174	86,63	86	Inkl. Schreibfehler; Auswertung der Orthographie siehe Tab.8
Total	5972	100	100	

Der Anteil der Kurzformen an der Gesamtmenge der Suchanfragen ist insofern interessant, als die Verwendung von Abkürzungen auf Suchende mit organisationspezifischen Kenntnissen verweist. In der Tab. 7 werden die verwendeten Kurzformen noch spezifiziert nach Akronymen, die die Geschäftsfelder der Organisation betreffen, hier der Hochschule, und auch im Hochschulumfeld bekannt sind, und nach organisationsinterner Sprachverwendung („Jargon“) unterteilt:

Tab. 7 Details Abkürzungen (interne Kenntnisse, relevant nur bei inhaltlichen Anfragen)

Art	Anfragen	% - Anteil gesamt		Bemerkungen
Bekannte Akronyme (B)	203	29,25	29	Im Hochschulumfeld bekannte Kurzformen wie <i>CAS</i> , <i>MAS</i> , <i>phd</i> , offizielle Kürzel der Institute und Zentren wie <i>IVM</i> , <i>ZPP</i>
Interner Jargon (J)	491	70,74	71	z.B. Kürzel der Studiengänge wie <i>jo</i> , Modulkürzel wie <i>mioek2</i> , interne Personenkürzel (4 Buchstaben) wie <i>anso</i> , intern bekannte Kürzel wie <i>vpn token</i>
Total	694	100	100	

Über zwei Drittel (71%) der Personen, die mit Kurzformen suchen, sind mit grosser Wahrscheinlichkeit Interne, das andere Drittel sind Personen mit guten Kenntnissen der Hochschule, wenn nicht auch Interne.

Orthographie und Rechtschreibung

Der Anteil der Suchanfragen mit Schreibfehlern wurde ermittelt, weil auf der Site-Suche der ZHAW kein Korrekturprogramm installiert ist und eine Suchanfrage mit Orthographiefehlern folglich zu einem Null-Treffer-Resultat führt. Bei einem höheren Anteil an fehlerhaften Suchanfragen müsste eine Lösung gefunden werden.

Tab. 8 Anzahl Schreibfehler (ohne Personenanfragen)

Art	Anfragen	% - Anteil gesamt		Bemerkungen
Unkenntnis Orthographie (O)	166	45,47	45	z.B. <i>flyght</i>
Tippfehler (T)	199	54,52	55	z.B. <i>flignt</i>
Total mit Schreibfehler	365	100		
Gesamttotal ohne Personen	5972	6,11	6	Anteil an der Gesamtmenge

Die sehr tiefe Zahl von 6% von Suchanfragen mit Schreibfehlern ist überraschend und muss sicher noch anhand weiterer Samples aus dem gleichen Korpus verifiziert werden. Da leider keine Vergleichszahlen mit Internet-Suchen vorliegen, können nur Vermutungen angestellt werden: entweder ist die grössere Aufmerksamkeit beim Schreiben der Suchanfrage eine weitere Charakteristik von Anfragen auf Site-Suchen, weil Schreibfehler hier eben i.R. zu Null-Treffern führen, oder aber sie ist ein Indiz für das durchschnittlich höhere Bildungsniveau der Nutzer einer Hochschulwebsite. Aufschlussreich sind weitere Details, dass nämlich die Unkenntnis der Orthographie bestimmter Wörter kein Einzelfall ist: *kursprogram* wurde in dieser Woche 18 mal gesucht, *somerkurse* 8 mal, *radiomoterator* und *moterator* insgesamt 10 mal oder *eintritt* 6 mal. Ebenso gibt es typische Vertipper wie *studnenplan* oder 4 mal der fast poetische Tippfehler *traumreservati-on*, oder jede Menge von Inversionen der Buchstabenreihenfolge wie *thearpie*, *blackba-ord* (könnte auch Unkenntnis der Schreibweise sein wie auch das häufige *blackbord*), *Volkartgeaebude* etc.

Lexik: Getrennt und Zusammenschreibung

Der Umgang mit Komposita bei der Formulierung von Suchanfragen ist in mehrerlei Hinsicht interessant. Komposita, die Verbindung von zwei und mehr Wörtern zu einem Wort mit erweiterter oder neuer Bedeutung ist ein charakteristisches Wortbildungsinstrument im Deutschen. Wenn in Datenbeständen neben dem zusammengesetzten Begriff auch die Bestandteile indexiert werden müssen, ist die Kompositazerlegung für das Information Retrieval u.U. eine Knacknuss. Zudem besteht im Deutschen die Möglichkeit, Komposita auch mit einem Bindestrich zu bilden, wenn ein Bestandteil als fremdsprachig empfunden oder die Zusammenfügung speziell markiert werden soll. So wurde zum Beispiel im vorliegenden Dokument bewusst die Schreibweise Site-Suche und Internet-Suche gewählt, um den Bedeutungsunterschied zwischen „Internet-Suche“ = Suchfunktion von Internetsuchmaschinen, und „Internetsuche“ = etwas im Internet suchen, zu markieren. Für den Nutzer ist unklar, wie Suchmaschinen mit Komposita und Bindestrich-Fügungen umgehen. Die Erfahrung mit Google zeigt zum Beispiel, dass dort der Bindestrich ignoriert und nach den beiden Begriffen getrennt gesucht wird. Will man bei Google den Komposita-Begriff erzwingen, muss er als Phrase mit Anführungszeichen eingegeben werden.

Es wird nun untersucht, wie viele feste Komposita, Bindestrich-Komposita und getrennt geschriebene bekannte Komposita im Sample (ohne Personennamen) vorkommen:

Tab. 9 Schreibweise von Komposita

Art	Anfragen	%-Anteil gesamt	Bemerkungen	
Komposita (K)	2149	35,98	37	Als Komposita wurden hier nur solche Fügungen gezählt, deren Bestandteile als eigenständige Nomen wahrgenommen werden können und die deshalb auch getrennt geschrieben wurden, wie z.B. <i>stundenplan</i> und <i>diplomarbeiten</i> . Hingegen wurden Begriffe, die nicht mehr als Komposita sondern als ein Wort empfunden werden wie <i>standort</i> , <i>oberstufe</i> hier nicht gezählt.
Zusammenfügungen mit Bindestrich (Z)	85	1,42		Wenig Fälle von festen Bezeichnungen wie <i>eduhelper</i> , sonst Kompositaformen wie <i>infoveranstaltung</i> , <i>absolventen-tag</i>
Getrenntschreibung möglicher Komposita (G)	72	1,20	2	Begriffe wie <i>info material</i> , <i>info mappe</i> , <i>copy center</i> , <i>anmeldung formular</i> , <i>assessment noten</i> , die mit Bindestrich oder sogar zusammengeschieden werden könnten
Getrenntschreibung fester Komposita (GK)	82	1,37		Komposita, die gängig zusammengeschieden werden, wie z. B. <i>aufnahme prüfung</i> , <i>einschreibe gebühr</i> , <i>arzt zeugnis</i> , <i>semester ferien</i>
Engl. Bezeichnungen	335	5,60	6	Bezeichnungen, die auf deutsch als Komposita geschrieben würden, wie <i>german course = Deutschkurs</i> , <i>flight traffic = Flugverkehr</i> , <i>credit points = Kreditpunkte</i> etc.
Total Kompositaformen	2723		45	
Gesamttotal alle ohne Personen	5972			100%

Die grosse Zahl von 45% von Komposita und Formen, bei denen die Kompositabildung ein Schreibproblem stellt, zeigt, dass der Umgang mit diesen Formen für Suchmaschinen in der deutschen Sprache eine zentrale retrievaltechnologische Frage ist. In der Gesamtauswertung wurden die englisch geschriebenen Suchanfragen, die häufig ein „Produkt“ betreffen (z.B. Weiterbildungsstudiengänge wie *mas chain supply management*, *cas care management*, oder *master in business administration*), speziell ausgezählt, weil entsprechende Bezeichnungen auf Deutsch meistens als Komposita (ein Wort) geschrieben würden. Immerhin 36% aller Wörter sind Komposita, die auch als ein graphematisches Wort, nämlich zusammen geschrieben wurden. Hingegen werden sehr wenige Bindestrich-Komposita (1,42 %) verwendet, was bestätigen würde, dass die Nutzer sie vermeiden, weil nicht klar ist, wie die Suchmaschine damit umgeht. Die gleiche Erklärung gilt vermutlich auch für die (sprachlich unbegründete) Getrenntschreibung von Komposita, hier wurde noch unterschieden zwischen Komposita, die auch mit Bindestrich geschrieben werden könnten (mögliche Komposita 1,20 %) und fester Komposita, die im normalen Sprachgebrauch zusammengeschieden werden wie *arzt zeugnis* (1,37%).

Diese Getrenntschreibung ist ein aufschlussreiches Beispiel, das noch genauer analysiert werden soll: das Informationsbedürfnis des oder der Suchenden war vermutlich eine administrative Auskunft darüber, ab wann ein Arzzeugnis erforderlich ist und bei Eingabe des gängigen Komposita *arztzeugnis* wären adäquate Treffer gefunden worden, falls eine

solche Informationsquelle existieren würde. Bei Getrennschreibung wird aber nach den Einzelbegriffen *arzt* und *zeugnis* gesucht, was im Hochschulkontext der ZHAW bestimmt die unerwünschten Ergebnisse bringt, nämlich Treffer zu Zeugnissen im Studienbereich und Treffer zu Arzt aus dem Fachbereich des Departementes Gesundheit. Die Auswertung zeigt übrigens, dass die gleiche Person noch wie folgt gesucht hat: *arzt-zeugnis*, „*arzt zeugnis*“ als Phrase markiert und schliesslich *arztzeugnis* – alle mit Null-Treffer-Ergebnis, da die entsprechende Informationsquelle im Intranet und nicht auf der Website zu finden ist. Dass die Nutzer generell wenig darüber wissen, wie eine Suchmaschine sucht, belegen auch die folgenden Beispiele eher unsinniger Schreibweisen:

qualitaets-management / eintritts-assessment	Bindestrichschreibung trotz Fugen-s, unklar, was mit der Bindestrichschreibung bezweckt wird
einfuehrungs pruefung / prüfungs ordnung	Getrennschreibung mit Fugen-s! Wenn nach den Einzelbegriffen gesucht werden soll, wäre Schreibweise ohne Fugen-s sinnvoller
studentischearbeiten / goehteinstitutpruefung	Grammatikalisch unmögliches Komposita, es gibt bestimmt keine Infoquelle dazu
"assistenten gehaelter" / "assistenten lohn"	Getrennt geschriebenes Komposita als Phrase markiert, keine Resultate, weil keine solche Phrase existiert, da unübliche Schreibweise

4.3 Inhaltliche Analyse

Themen

Eine inhaltlich-thematische Einteilung der Suchanfragen kann nur qualitativ und mit entsprechendem Aufwand geleistet werden, dies gilt auch für entsprechende Untersuchungen bei Logfile-Analysen von Internet-Suchen (Lewandowski 2006). Da die Menge der Infoquellen bei Site-Suchen begrenzt ist, lassen sich hier aber die möglichen Themenfelder besser definieren und das inhaltliche Clustering feiner abstimmen. Grundsätzlich betreffen auf einer Site-Suche die Suchanfragen inhaltlich das Themenspektrum der Organisation, hier im Fall der ZHAW die Fachbereiche der acht Departemente: Architektur, Gestaltung und Bauingenieurwesen; Gesundheit; Angewandte Linguistik; Life Sciences & Facility Management; Angewandte Psychologie; Soziale Arbeit; Engineering sowie Management & Law, mit all ihren entsprechenden Unterbereichen und Subthemen. Für die vorliegende Untersuchung interessierte jedoch nicht eine quantitative Zuordnung der Suchanfragen zu diesen Themenbereichen, sondern die Frage, ob sich die Suchanfragen nach inhaltlichen Mustern gliedern lassen, die allenfalls für die Site-Suche typisch sind. Es wurde dabei iterativ vorgegangen, nach einer ersten Durchsicht aller 2469 unterschiedlichen inhaltlichen Suchanfragen fiel auf, dass es sehr viele Anfragen nach administrativen oder organisationsspezifischen Belangen gibt, mehr sogar als fachthematische. Folglich wurden administrative Anfragen und fachthematische Anfragen genauer definiert und in einem ersten Durchgang wurde das ganze Sample ohne Eigennamen, d.h. auch Abkürzungen, Fehlerhaftes und Unklares, diesen zwei Gruppen zugeteilt. In einem nächsten Schritt wurden beide Kategorien noch feiner unterteilt, wobei jede weitere Differenzierung nur von Personen mit guten Kenntnissen der Organisation, ihren Themen und Produkten geleistet werden kann.

Tab. 10 administrative und fachliche Anfragen

Art	Anfragen	%Anteil gesamt		Bemerkungen
Administrativ	3694	61,85	62	Anfragen nach Prüfungen, Tests, Kursprogrammen, Stundenplänen etc.
Fachthemen	2156	36,10	36	(inhaltliche) Anfragen nach Studiengängen, Studienabschlüssen, Diplomen, Fachbegriffen etc.
Nicht zuordbar	122	2,04	2	Rest aus den inhaltlichen Gruppen „Fehler“ und „Unklares“
Gesamttotal ohne Personen	5972	100		= 2469 unterschiedliche Anfragen

Die Anzahl administrativer Anfragen ist mit fast Zweidritteln (62%) aller Suchanfragen überraschend hoch, was ein weiteres Indiz dafür sein könnte, dass viele Interne und viele Studierende die Site-Suche benutzen. Wiederum aufgrund von Annahmen wurde eine weitere Unterteilung der administrativen Anfragen nach studienrelevanten Belangen, nach Anlässen (Termine, Orte) und nach Räumen und Standorten vorgenommen.

Tab. 11 Details zu administrativen Anfragen (a)

Art	Anfragen	%Anteil gesamt		Bemerkungen
Studienrelevante Belange	1696	45,91	46	Semestertermine, Prüfungsdaten, Prüfungsordnung, Studienrichtlinien, Jahrestermine, Stundenpläne
Anlässe	82	2,21	2	Davon betreffen 16 Anfragen nur intern bekannte Anlässe, würden im Intranet gefunden
Räume, Standorte	191	5,17	5	Auch Fragen nach Adressen, Gebäuden
Rest	1725	46,69	47	Anfragen wie z.B. <i>mietwohnung, kadermitglieder, interessenliste, dozentin, sponsor, auftakt, baustelle, bueromaterial, fundgegenstaende, beschwerde, katalog, beitraege, auslaender, demo englisch, berechnung leasing, kognitver test</i> etc. die sich nur noch mit grossem Aufwand weiter clustern lassen
Total administrative Anfragen	3694	100		

Auch innerhalb der administrativen Suchanfragen betrifft fast die Hälfte studienrelevante Belange, die mit grosser Wahrscheinlichkeit von Studierenden gestellt wurden. Ein erster Überblick über die andere Hälfte zeigt als relevante Häufung nur noch die Frage nach lokalen Informationen (Räume, Gebäude, Standorte, Adressen), wobei die Detailanalyse hier zeigt, dass häufig mit dem internen Sprachgebrauch gesucht wurde (mehrmals *Halle 180*, der Standort des Departementes Architektur, der aber offiziell nirgends mit dieser Bezeichnung aufgeführt ist), was wiederum belegt, dass diese Suchanfragen von Interne resp. Studierenden stammen. Der Prozentsatz der Suchanfragen nach Anlässen liegt mit 2% sehr tief, das kann sich jedoch je nach Zeitpunkt ändern, wenn zum Beispiel ein grösserer Anlass der Hochschule im Untersuchungszeitraum liegt.

Der grosse Prozentsatz an administrativen Anfragen könnte spezifisch für die untersuchte Site einer Hochschule sein. Hauptsächlich Studierende als halbinterne Nutzer, aber

auch einige Assistierende und Dozierende, suchen administrative Informationen oft auf der Website und nicht im Intranet aus Gründen von Bequemlichkeit (Intranet verlangt Login) und Nicht-zur-Kennntnis-nehmen von administrativen Inhalten (auf dem Intranet). Viele Hochschulen verfügen auch nicht über Intranets als geschlossene Bereiche, wie sie in Unternehmen existieren. Es ist deshalb zu vermuten, dass dieselben Anteile bei Suchanfragen sich auch bei andern Hochschul-Websites nachweisen lassen, dass jedoch bei Websites von Unternehmen der Prozentsatz an administrativen Anfragen viel geringer sein dürfte.

Tab. 12 Details zu fachthematischen Anfragen (f)

Art	Anfragen	% - Anteil gesamt		Bemerkungen
„Studienprodukte“	621	28,80	29	Mit Namen benannte konkrete Kurse und Abschlüsse (<i>Master, MAS, CAS, DAS</i> etc.)
Fachgebiete der ZHAW	491	22,77	23	ZHAW-Bezeichnungen von (Studien-)Fachgebieten, Studiengängen, Kursen und Berufen wie Wirtschaftsinformatiker, Risikomanagement, Elektrotechnik oder Pflege etc.
Detailthemen	1044	48,42	48	Alle fachthematischen Anfragen, die eine Detailthematik betreffen, wie <i>gartengestaltung, weichkäse, häusliche gewalt, solar, sparquote, coffee, air traffic</i> etc.
Total fachthematische Anfragen	2156	100		

Auch bei den fachthematischen Anfragen betrifft über die Hälfte (29% + 23%) studienrelevante Anfragen. Die Kategorien Fachgebiete der ZHAW und Detailthemen wurden wiederum so unterschieden, dass die Suchanfragen in der ersten Kategorie auf Kenntnisse der ZHAW deuten lassen und Anfragen der zweiten Kategorie eher auf ein Herantasten im Stile von „ausprobieren, ob es Inhalte dazu gibt“. Nach subjektiver Einschätzung stammen viele dieser Begriffe aus Projekten der Hochschule, über die in den Medien berichtet wurde. Um diese Vermutung zu erhärten, müssten allerdings die Suchanfragen mit der Auswertung der Medienberichterstattung korreliert werden.

Werden nun alle Anfragen zusammengefasst, die irgendwie studienrelevante Themen betreffen, ergibt sich folgende Verteilung:

Tab. 13 Gesamtanteil studienthematische Anfragen

Art	Anfragen	Anteil gesamt		Bemerkungen
„Studienprodukte“ (f)	621	2808	48%	Studienthemen
Fachgebiete der ZHAW (f)	491			
Studienrelevante Belange (a)	1696			
Detailthemen (f)	1044	1044	18%	Begriffe aus Projekten, Forschung
Anlässe (a)	82	1998	34%	Administrative, organisatorische Themen
Räume, Standorte (a)	191			
Rest (a)	1725			
Total inhaltliche Anfragen ohne die nicht Zuordbaren, vgl. Tab.10	5850		100%	

Dass praktisch die Hälfte aller Suchanfragen (ohne die Suche nach Personen) studienrelevante Themen betrifft, lässt verschiedene Folgerungen zu: mit grosser Wahrscheinlichkeit kann bei diesen Anfragen auf Studierende und Studieninteressierte geschlossen werden. Dies wäre ein Indiz für die Richtigkeit der Hypothese, dass diese jüngere Nutzergruppe, die Internetgeneration, als Suchstrategie auf einer Website prioritär die Site-Suche benutzt und nicht mit Navigieren sucht. Es lässt sich daraus aber nicht folgern, dass die Hälfte der Nutzer der Website zu dieser Gruppe gehören, denn ältere Nutzer und solche aus dem wissenschaftlichen Umfeld, die auf Websites vermutlich eher strukturiert über Navigieren suchen, werden bei der Auswertung der Suchanfragen ja nicht erfasst.

Anfragetypen

Die in der Retrievalliteratur viel zitierte (u.a. Braschler et al. 2006 / Lewandowski 2005, 2006 / Stock 2007) Unterteilung von Broder in drei Anfragetypen bei Internet-Suchen, nämlich *navigationsorientiert* (Suche nach einer bestimmten Website als Ausgangspunkt für Weitersuchen mit Navigieren), *informationsorientiert* (Suche nach allen möglichen Informationen zu einem bestimmten Stichwort in einem Datenbestand, klassische Information Retrieval-Aufgabe) und *transaktionsorientiert* (Suchen einer Quelle zwecks einer Transaktion wie Download, Kauf etc.) bezieht sich auf Internet-Suchen. Aufgrund der Unterschiede zwischen Internet-Suchen und Site-Suchen ist zu vermuten, dass diese Kategorien von Anfragetypen nur bedingt anwendbar resp. differenziert und angepasst werden müssen.

Broder analysierte 2002 rund 400 Suchanfragen an die Internet-Suche Alta Vista einerseits aus Logfileanalysen, andererseits anhand einer Nutzerbefragung, gemäss seinen Anfragetypen navigationsorientiert (n), informationsorientiert (i) und transaktionsorientiert (t). Er fand folgende Verteilung: n = 20-25%, i = 39-48%, t = 22-36% (die Spannweite ergibt sich aus den beiden unterschiedlichen Untersuchungen). Um die Aktualität dieser Verteilung zu verifizieren, wertete Lewandowski 2006 in einer ähnlich angelegten Untersuchung 1500 Anfragen an die deutschen Suchmaschinen Fireball, Metager und Seekport aus und kam zu folgenden Durchschnittswerten: n = 40%, i = 42-47%, t = 11-18% (Lewandowski 2006, 2009:3f). Für die Verschiebung gibt es keine eindeutige Erklärung, sie könnte durch den Zeitunterschied oder durch andere Internetnutzungen der deutschen Nutzer begründet sein.

Abschliessend soll also noch untersucht werden, ob die erwähnten Anfragetypen bei Internet-Suchen auch eine brauchbare Kategorisierung der unterschiedlichen Anfragen bei einer Site-Suche darstellen. Eine erste Grobanalyse zeigte schnell, dass die Einteilung in diese Kategorien wenig Erkenntnisgewinn gibt, viel relevanter scheint das Kriterium, ob die Suchanfrage präzise ist und mit genau einem Treffer beantwortet werden kann, oder ob sie unspezifisch ist und verschiedene Informationen zur Beantwortung benötigt. Dies entspricht den zwei Typen von Informationsbedürfnissen: *Suche nach etwas Bekanntem* und *Suche nach etwas Unbekanntem*. Eine ähnliche Unterscheidung wird von Stock im Information Retrieval auch als *Concrete Information Need* und als *Problem Oriented Information Need* bezeichnet (Stock 2007:51) und wie folgt charakterisiert: beim *Concrete Information Need* ist das thematische Feld klar abgesteckt und die Suchformu-

lierung durch exakte Terme ausgedrückt, eine Fakteninformation reicht i.R. aus, um das Informationsbedürfnis zu befriedigen. Beim *Problem Oriented Information Need* hingegen sind die thematischen Grenzen nicht klar bestimmbar und die Formulierung lässt mehrere terminologische Varianten zu, es müssen i.R. mehrere Informationsquellen beschafft werden und ob das Informationsbedürfnis befriedigt wird, bleibt offen. Bei einer Site-Suche mit begrenzten Informationsquellen ist es für eine Person mit guten Kenntnissen der Website möglich, Anfragen danach zu charakterisieren, ob sie mit einem Treffer beantwortet werden könnten, allerdings setzt dies eine aufwändige qualitative Analyse voraus.

Wir bezeichnen das *Concrete Information Need* oder die Suche nach dem Bekannten als *präzise Anfrage* und das *Problem Oriented Information Need* oder die Suche nach etwas Unbekanntem als *unspezifische Anfrage*. Um die Suchanfragen auszuwerten, wurden folgende begründbare Annahmen getroffen:

1. alle Anfragen nach Personen sind präzise Anfragen, da an der ZHAW der Site-Suche als Infoquelle eine Personendatenbank (Porträt Datenbank) zur Verfügung steht. Nach Eingabe eines Eigennamens liefert die Site-Suche als ersten Treffer den Datenbankeintrag dieser Person. Dieses Porträt kann diverse Links auf Publikationen, Projekte und Website dieser Person enthalten, es ist möglich, von dort aus weiterzunavigieren. Die Suchanfrage nach einer Person kann hier also auch als navigationsorientiert betrachtet werden.
2. Alle Anfragen, die direkt (*alle Studiengänge*) oder indirekt (*Weiterbildungen*) nach Übersichtsseiten suchen, können ebenfalls als präzise und navigationsorientiert bezeichnet werden
3. Alle Anfragen mit genauen Bezeichnungen von Angeboten und Kursen sowie Anfragen mit Termen, die der offiziellen ZHAW-Terminologie entsprechen, sind präzise Anfragen, aber nicht unbedingt navigationsorientiert.
4. Viele Akronyme und alle Kurzformen, die als Jargon bezeichnet wurden, stellen präzise Anfragen dar.
5. weiter wurden alle Anfragen markiert, die nach Formularen, Richtlinien, Anweisungen, Broschüren etc. suchten, sie werden als präzise Anfragen und als transaktionsorientiert betrachtet, da es wahrscheinlich ist, dass der Download dieser Dokumente erwünscht wird.

Sämtliche Suchanfragen wurden gemäss den Punkten 2 - 5 ausgewertet, wobei ein wichtiges Kriterium für die Zuordnung zu „präzise Anfrage“ das mögliche Vorhandensein einer Informationsquelle war, die diese Anfrage beantworten könnte.

Tab. 14 Präzise und unspezifische Anfragen

Art	Anfragen	%-Anteil an präzisen	Bemerkungen
Eigennamen	2072	31,37	31 präzise + navigationsorientiert (1. Treffer Eintrag Porträt Datenbank mit u. U. vielen Links)
„Studienprodukte“ (f)	621	9,40	9 präzise (Kenntnis der Produkte)
Fachgebiete der ZHAW (f)	491	7,43	8 Präzise (Kenntnis der Fachgebiete und ZHAW-Terminologie)
Studienrelevante Belange (a)	1696	25,68	26 Präzise (Kenntnis der Terminologie)
Anlässe, Standorte (a)	273	4,13	4 Präzise (Kenntnis der Bezeichnungen und Orte)
Weitere präzise Anfragen aus „Detailthemen“(f)	305	4,61	5 Die Abgrenzung zu unspezifischen Anfragen war nicht immer eindeutig, Zweifelsfälle wurden als unspezifisch eingeteilt
Weitere präzise Anfragen aus „Rest“(a)	1145	17,34	17 Auch präzise Anfragen, zu denen keine Infoquellen auf der Website jedoch im Intranet oder extern vorhanden sind, wurden hier erfasst, wie z.B. <i>sbb</i> (Schweiz. Bundesbahnen, extern) oder <i>individuelle Leistungsvereinbarung</i> (Intranet)
Total präzise Anfragen	6603	100	
Gesamttotal Anfragen	8044	82,08	82

Die präzisen Anfragen, die den Anfragetypen „navigationsorientiert“ + „transaktionsorientiert“ oder der *Suche nach etwas Bekanntem*, auch *Concrete Information Need*, entsprechen, machen mit vier Fünfteln der Gesamtmenge aller Anfragen einen sehr hohen Anteil aus. Auch wenn noch aufgrund der qualitativen Bewertung mit einer Beurteilungstoleranz von ca. 10% gerechnet werden muss, bleibt der Anteil sehr gross, was als wichtigstes Ergebnis der vorliegenden Untersuchung betrachtet werden kann. Die Untersuchung von Braschler et al. (2006:29) ging bei Suchanfragen auf Websites von einem Verhältnis von 2:1 (zwei Drittel navigations- + faktenorientiert [ersetzen dort transaktionsorientiert], ein Drittel informationsorientiert) aus, gemäss dem vorliegenden Ergebnis wäre das Verhältnis auf einer Hochschulwebsite sogar etwa 4:1. Selbstverständlich muss dieses Resultat erstens mit weiteren Samples aus dem gleichen Korpus, zweitens mit Suchanfragen aus der Site-Suche einer andern Hochschule und drittens einer Unternehmens-Website verifiziert werden.

„Navigationsorientiert“ und „transaktionsorientiert“ sind insofern gleich, als das Informationsbedürfnis mit einem richtigen Treffer befriedigt werden kann, sie unterscheiden sich aber darin, dass der Nutzer eine andere Art von Informationsquelle erwartet: bei ersteren Seiten mit Links zum Weiternavigieren, bei letzterem eine Downloadmöglichkeit. Es wurde deshalb untersucht, bei wie vielen Anfragen explizit ein Dokument gesucht wird, ausgehend von der Annahme, dass in einer Suchanfrage wie *dokument* die Präsupposition „downloaden“ oder „ausdrucken“ enthalten ist.

Tab. 15 Anzahl transaktionsorientierter Anfragen

Art	Anfragen	% - Anteil gesamt		Bemerkungen
Administrativ	106	88,33	88	Nur eindeutige Anfragen nach Dokumenten wie <i>Reglement, Formular, Vorlage, Broschüre, Infomappe</i> oder <i>pdf Studiengang</i> wurden erfasst, aber nicht <i>Lagepläne, Stundenplan, Semestertermine, Agenda, Kalender</i> etc., bei denen vermutlich ebenfalls oft ein Download oder Ausdruck erwünscht wird.
Fachthemen	14	11,66	12	z.B. Suche nach <i>Lehrmittel, Flugplan, Lebensmitteltest</i>
Total transaktionsorientiert	120	100		
Gesamttotal zuordbar + ohne Personen	5850		2	

Wie zu erwarten war, sind transaktionsorientierte Anfragen mehrheitlich nur bei administrativen Suchanfragen zu finden. Die Gesamtzahl der transaktionsorientierten Anfragen ist mit 2% so tief, weil nur solche Suchanfragen erfasst wurden, bei denen eindeutig auf einen Download-Wunsch geschlossen werden kann.

Für das in der Suchapplikation implementierte Information Retrieval ist die Anzahl Terme pro Anfrage wichtig. Bei 4.1. wurde festgestellt, dass der Prozentsatz an Ein-Wort-Anfragen mit 70% (ohne Personenanfragen) viel höher ist als bei Internet-Suchen, deshalb wurde bei den präzisen Inhaltsanfragen (ohne Personensuche) noch die Verteilung der Wortzahl pro Anfrage untersucht:

Tab. 16 Anzahl Terme pro Anfrage bei präzisen Anfragen (vgl. Tab. 1)

Wörter pro Anfrage	% - Anteil fachthematische präzise Anfragen	% - Anteil administrative präzise Anfragen	% - Anteil gesamt ohne Personensuche	Vergl. Internet-Suche Ø % D	Vergl. Internet-Suche Ø % E
1	46	71	70	50	15
2	34	22	20	?	32
3	12	5	7	?	27
4 + mehr	8	2	2	?	15

Das Resultat ist in zweierlei Hinsicht bemerkenswert: erstens entspricht die Verteilung bei den präzisen administrativen Anfragen (71 – 22 – 5 -2) ziemlich genau den Prozentsätzen des ganzen Samples (70 – 20 – 7 – 2), was die Vermutung bestätigt, dass die Nutzer auf Websites bei administrativen Themen prioritär Bekanntes suchen, da Ein-Wort-Anfragen ein Ausdruck dafür sind, dass der Suchende das Gefühl hat zu wissen, was er sucht, oder sich sicher ist, dass das Gesuchte vorhanden ist. Diese Tatsache verschärft jedoch noch die ganze Terminologie- und Benennungsproblematik, denn bei ungenauer Bezeichnung erhält der Suchende dann häufig Null-Treffer-Ergebnisse. Erstaunlich ist zweitens, dass bei den fachthematischen Suchanfragen der Anteil an Mehr-Wort-Anfragen viel höher liegt und mit 54% ungefähr den Werten für deutsche Anfragen bei

Internet-Suchen entspricht. Daraus kann gefolgert werden, dass die fachthematische Anfrage auf einer Website vom Informationsbedürfnis her am ehesten mit der Art der Suche im Internet vergleichbar ist, auch bei präzisen Anfragen, denn nur diese wurden ja berücksichtigt, bestehen fachthematische Begriffe häufiger aus Mehrwortanfragen.

Null-Treffer-Anfragen

Aus der Sicht der Kommunikation (Website-Betreiberin) ist die Analyse der Null-Treffer relevant, da die Resultate direkte Hinweise auf das Optimierungspotenzial der Website liefern können. Abschliessend werden deshalb noch die Suchanfragen ausgewertet, die zu Null-Treffer-Ergebnissen führten. Darin enthalten sind die bereits erwähnten Anfragegruppen Orthographiefehler, Unklares und fehlerhafte Operatoren, diese wurden hier natürlich nicht berücksichtigt. Für die vorliegende Analyse wurde ermittelt, wie viele der Null-Treffer-Anfragen signifikante Informationsbedürfnisse ausdrücken, für die Massnahmen getroffen werden müssen, damit die Site-Suche Treffer liefert.

Tab. 17 Null-Treffer-Anfragen

Art	Anfragen	Verschiedene	%-Anteil gesamt	Bemerkungen
administrative Anfragen	93	35	8	Oft andere Terminologie wie <i>fachmatura</i> anstatt <i>berufsmatura</i> , <i>kantine</i> anstatt <i>mensa</i> , oder Inhalte, die klar nur im Intranet vorhanden sind wie <i>einzelentschädigungsformular</i> , <i>personalienblatt</i> , <i>sabbatical</i> etc.
Studienrelevante Anfragen	139	45	12	Anfragen wie <i>pruefungsraueme</i> , <i>schulfrei</i> , <i>schultage</i> , oder <i>studentaccount</i> , <i>matrikelnummer</i> , <i>pruefungsdaten</i> (alle Intranet)
Fachthematische Anfragen	112	36	10	Anfragen wie <i>zwetschgen</i> , <i>brotrad</i> , <i>islam</i> , <i>pflanzenraritaeten</i> , <i>polymertechnologie</i> , <i>notfallpsychologe</i> etc.
Total korrekte Anfragen	344	116	31	<i>Ihre Relevanz und die Informationsbedürfnisse müssten aus Sicht der Kommunikation untersucht werden</i>
Kurzformen	119	31	11	Meist Modulkürzel wie <i>wi08b</i> , <i>pf07pie</i> oder Räume wie <i>tb265</i>
Fehlerhaftes, inkl. Namen	662		59	
Total Null-Treffer	1125	396	14	
Gesamttotal	8044	2469		

Im Untersuchungszeitraum hatten nur 14% aller Anfragen zu einem Null-Treffer-Ergebnis geführt, was als sehr gutes Resultat bezeichnet werden kann. Insgesamt fand die Site-Suche für 1125 Anfragen keine Treffer, davon 396 verschiedene Anfragen. Ohne die Anfragegruppen Orthographiefehler, Unklares, fehlerhafte Operatoren und Abkürzungen bleiben noch 344 korrekt geschriebene und sinnvolle Anfragen (davon 116 verschiedene), was ein Drittel aller Null-Treffer-Anfragen ausmacht. Diese Anfragen müssen genau-

er untersucht werden, ob sie relevante Informationsbedürfnisse ausdrücken, und wenn ja, ob es dazu a) Informationsquellen gibt, oder b) allenfalls Informationen bereitgestellt werden müssten. Falls passende Informationsquellen vorhanden sind, kann dort eventuell die Terminologie so angepasst oder ergänzt werden, dass die Suchmaschine sie findet, oder bei der Suchapplikation Relevancy der ZHAW können Informationsbedürfnisse ohne Treffer manuell mit Suchtipps versehen werden, die direkt auf die richtige Seite verlinken. Da im Nutzungsmonitor auch die Häufigkeit, mit der dieselbe Anfrage gestellt wird, ausgewertet ist, kann bei häufigen Suchanfragen mit Null Treffern sofort gehandelt werden. Zum Beispiel führte die Anfrage *jobs* auf der ZHAW-Website zu Null Treffern und ist immer eine häufige Anfrage. Mit dem Suchtipp und Link auf *Offene Stellen* können Suchende nun mit einem Klick auf die richtige Seite verwiesen werden.

5. Fazit und Ausblick

5.1 Unterschiede zu Internet-Suchen

Eine der zentralen Fragen der vorliegenden Studie war, ob sich die Merkmale des Suchverhaltens, die bei Internet-Suchmaschinen festgestellt wurden, auch bei einer Site-Suche nachweisen lassen. Die Frage kann mit teils-teils beantwortet werden. Keine Unterschiede zeigen sich bei der Verwendung von Operatoren, sie lag unter 3% sowohl bei den Internet-Suchen wie auch bei der Site-Suche, wovon im Internet ca. die Hälfte und auf der Website ca. ein Drittel falsch verwendet wurden. Die heute sehr breite Nutzergruppe, die die Suchfunktion benutzt, hat die Verwendung von Operatoren und professionellere Suchstrategien nie gelernt, es wird generell intuitiv gesucht. Hingegen brachte der Vergleich der Suchanfragen auf Internet-Suchen und auf dieser Site-Suche sowie die linguistischen und inhaltlichen Analysen folgende signifikante Unterschiede zu Tage:

- 1) ein markant höherer Anteil an Ein-Wort-Anfragen bei der Site-Suche (70% zu ca. 50%), deutet auf Unkenntnis spezieller Suchstrategien
- 2) ein Viertel der Suchanfragen betreffen Personen (Angehörige der Organisation)
- 3) ein sehr hoher Anteil an präzisen Anfragen (navigationsorientiert + transaktionsorientiert oder Suche nach etwas Bekanntem resp. Concrete Information Need) auf der Hochschulwebsite (80% zu ca. 55% bei Internet-Suchen).
- 4) 60-80% Nutzer mit Kenntnissen der Organisation (Verwendung von Organisations-Jargon, genauer Terminologie und vor allem administrativen Begriffen)

Erklärungen dafür können nach dieser Pilotstudie erst als Vermutungen formuliert werden, sie müssen mit weiteren Samples aus dem gleichen Korpus und mit Untersuchungen von Korpora aus Site-Suchen von andern Websites verifiziert werden. Der hohe Anteil an Ein-Wort-Anfragen (1) korreliert auf jeden Fall mit dem ebenfalls hohen Prozentsatz an präzisen Anfragen (3) bei der Site-Suche, auch wenn dieselben Werte bei andern Websites etwas tiefer liegen können. Von allen Nutzern, die die Suchfunktion auf einer Website benutzen, ist der Anteil der Nutzer mit Kenntnissen der Organisation auf jeden Fall bedeutsam. Suchende mit Kenntnissen der Organisation haben ganz bestimmte Vorstellungen, was sie suchen, sie sind überzeugt, dass der Sachverhalt vorhanden sein muss, kennen jedoch nicht immer die dafür verwendete Terminologie. Typisch für Conc-

rete Information Needs oder die Suche nach etwas Bekanntem (präzise Anfrage) ist folglich, dass die Erwartungen des Suchenden an die Site-Suche hoch sind.

5.2 Konkrete Informationsbedürfnisse und hohe Erwartungen

Nun sind präzise Anfragen (Concrete Information Needs) auf Site-Suchen aber insofern anspruchsvoller als auf Internet-Suchen, weil präzise die Informationsquelle, die die Anfrage beantwortet, als Treffer geliefert werden sollte. Dazu müssen bei einer Site-Suche retrievaltechnische Rankings der Suchmaschine, website-spezifische Konfigurationen der Suche und die Aufbereitung der durchsuchbaren Informationsquellen aufeinander angepasst sein und zusammenspielen (vgl. auch Braschler et al. 2006:7), was um einiges komplexer ist, als aus einer praktisch unbegrenzten Informationsmenge im Internet passende Treffer zu finden. Dazu kommt wie bereits erwähnt, dass der Suchende auf Websites höhere Ansprüche an die Treffergenauigkeit stellt als bei der Suche auf einer Internet-Suchmaschine, obwohl er ähnliche bis gleiche Suchstrategien anwendet. Da bei Internet-Suchen sowohl bei präzisen wie auch bei unspezifischen Anfragen riesige Treffermengen geliefert werden, findet der Nutzer immer etwas und ist damit zufrieden (Schmidt-Mänz, Bomhardt 2005), weil er gar nicht weiss, welche andern noch besseren Treffer existieren. Bei der Suche auf einer Website hingegen, insbesondere bei internen und halbinternen Nutzern, hat der Suchende oft Kenntnis des Sachverhalts resp. der Existenz der Informationsquelle und erwartet den oder die richtigen Treffer. Kennt er nun die genaue Bezeichnung für den Sachverhalt nicht, die die Organisation braucht oder die auf der Website verwendet wird, bekommt er schnell Null-Treffer, was bei einer Internet-Suche praktisch nie der Fall ist. Die Analyse der Nulltreffer-Anfragen aus dem Nutzungsmonitor belegt klar das Benennungsproblem, wenn mehrere ähnliche Anfragen zeigen, dass Nutzer unterschiedliche Formen (Zusammenschreibung, Getrennschreibung), den Einsatz von Operatoren, mit Ein- und Mehrwortanfragen oder mit Oberbegriffen resp. Synonymen ihr Formulierungsglück versuchen. Das Suchdilemma der Nutzer bei Site-Suchen kann somit präzisiert werden: die Nutzer wissen meist sehr genau, was sie suchen, resp. finden möchten, kennen aber oft die auf der Website gebräuchliche Bezeichnung für die Sache resp. die Informationsquelle nicht.

5.3 Kaum Suchstrategien

Der hohe Anteil an Ein-Wort-Anfragen deutet auch darauf hin, dass die Nutzer für präzise Anfragen auf Websites über keine speziellen Suchstrategien verfügen, wie zum Beispiel ihre Anfrage mit weiteren Begriffen zu präzisieren. Die Anzahl verschiedener Varianten von Suchanfragen zu offensichtlich demselben Informationsproblem (siehe Beispiel *arztzeugnis*, solches Sucherverhalten wurde hier aber noch nicht systematisch ausgewertet,) zeigt, dass die Nutzer bei unbefriedigenden Treffern oder Null-Treffer-Ergebnissen als einzige Strategie mit „Versuch und Irrtum“ einfach weitere Ein-Wort-Schreibvarianten eingeben oder im Fall von Komposita diese in die Elemente aufteilen mit, wie wir gesehen haben, zum Teil absurder Schreibweise (z.B. Getrennschreibung aber mit Fugen-s). Ein Grund für die Präferenz von Ein-Wort-Anfragen auf der Site-Suche könnten aber auch die Erfahrungen der Nutzer mit Internet-Suchen sein. Internet-Suchen liefern meist

bei Mehr-Wort-Anfragen zu jedem Term einzelne Treffer, was schon auf den ersten Ergebnisseiten zu vielen unerwünschten Treffern führt.

Ein weiterer, vermutlich gewichtiger Grund könnte sein, dass der grösste Teil der Nutzer dieser Site-Suche trotz Kenntnissen der Organisation und trotz subjektiver Gewissheit, dass die gesuchte Information irgendwo auf der Website existieren muss, wenig bis keinerlei Vorstellungen über den organisatorischen Kontext hat, zu dem der gesuchte Sachverhalt gehört, weshalb er weder Kontext- (z.B. präzisierendes Verb, Adjektiv, Nomen) noch Kontext-Begriffe (z.B. den Fachbereich) eingibt. Die Detailanalyse der Mehrwort-Anfragen lässt sogar den Verdacht aufkommen, dass viele Nutzer der Site-Suche über keinerlei Vorstellungen verfügen, wie die Suchmaschine sucht. Wer *Frau Müller* als Suchbegriff eingibt, hat sich wohl nicht überlegt, dass kaum irgendwo auf der Website „Frau Müller“ stehen wird. Der Einsatz interaktiver Möglichkeiten zur Präzisierung der Anfrage ist deshalb nur sinnvoll, wenn die vorgeschlagenen Ergänzungen dem Suchenden Zusatzwissen liefern, das er nicht hatte. Die Nähe zwischen Nutzergruppen und Website dürfte die Bereitschaft des Suchenden zum Dialog verstärken. Die Entwicklung von interaktiven Elementen, die die Sucheingabe ergänzen, bedingt jedoch immer manuelle Analysearbeit und kommunikative Kenntnisse. Diesen Vorteil haben die Betreiber der Site-Suche, da sie die durchsuchten Inhalte theoretisch kennen, was ihnen eher ermöglicht, bei unpräzisen Suchanfragen die Informationsbedürfnisse zu erkennen und sie sinnvollen Informationsquellen zuzuordnen. Das grundsätzliche Dilemma bei Site-Suchen ist jedoch, dass das Risiko von Null-Treffer-Ergebnissen steigt, je mehr man die Suchenden präzisiert, d.h. die Anzahl der Informationsquellen einschränken lässt.

5.4 Optimierungspotenzial für die Kommunikation

Abschliessend kann skizziert werden, wie die sprachlichen Spuren, die die Suchenden bei der Formulierung ihrer Informationsbedürfnisse in der Site-Suche hinterlassen, von der Kommunikation genutzt werden können, um die Website zu optimieren. Die Auflistung der Null-Treffer-Anfragen, die der Nutzungsmonitor von Relevancy liefert, ist eine erste und einfache Möglichkeit, Informationsbedürfnisse zu erkennen, die die Site-Suche nicht befriedigen konnte. Die Null-Treffer-Suchanfragen müssten regelmässig ausgewertet werden, indem zuerst alle fehlerhaften Anfragen eliminiert und anschliessend die korrekten Suchbegriffe genauer bezüglich Relevanz des Informationsbedürfnisses aus Sicht der Kommunikation analysiert werden, unterstützt durch die Angabe der Häufigkeit, mit der eine Null-Treffer-Anfrage im untersuchten Zeitraum gestellt wurde. Ebenfalls regelmässig müssten den häufig gestellten und als relevant erachteten Null-Treffer-Anfragen durch manuelle Analyse hypothetische Informationsbedürfnisse zugeordnet, resp. definiert werden, welche vorhandenen oder allenfalls noch zu schaffenden Informationsquellen auf der Website eine solche Anfrage optimal beantworten würden.

Weiter müssten bei allen häufig gestellten Anfragen immer wieder mal stichprobenartig die ausgegebenen Treffer inklusive ihr Ranking auf ihre kommunikative Relevanz bezüglich Nutzerbedürfnisse und dafür optimale Quellen überprüft werden. Zum Beispiel eine der häufigsten Anfragen nach „offenen Stellen“ sollte als ersten Treffer die auf der Website existierende Übersichtsseite über sämtliche Stellenangebote der Organisation (optima-

le Quelle, hohe kommunikative Relevanz) bringen und nicht irgendwelche Unterseiten (wegen retrievaltechnischer Relevanz). Bei nicht optimalen oder Null-Treffer-Ergebnissen müssen die Metadaten, Seitentitel oder die Terminologie der Seite angepasst werden, die die optimale Informationsquelle wäre, damit die Site-Suche sie findet und richtig indexiert. Eine weitere Möglichkeit, die der Nutzungsmonitor von Relevancy bietet, ist, häufig gestellte und wichtige Anfragen einem Informationsbedürfnis zuzuordnen (Clustering) und mit einem entsprechenden Suchtipp („Promolink“) zu versehen, wie das oben beim Beispiel *jobs* erläutert wurde. Der Hauptunterschied zwischen Internet-Suchen und Site-Suchen besteht darin, dass bei letzteren die Qualität der Ergebnisse gut überprüft und beurteilt werden kann, weil die Informationsquellen bekannt sind. Die Trefferqualität kann auch entsprechend verbessert werden, wenn die Informationsquellen für die Suche optimiert werden können.

Dies alles zeigt, dass das erwähnte wünschenswerte Zusammenspiel zwischen retrieval-technologischem Ranking, website-spezifischen Konfigurationen der Suche und entsprechender Aufbereitung der durchsuchbaren Informationsquellen komplex und aufwändig ist. Wieviel Aufwand sich seitens der Online-Kommunikation für eine gute Erschließung der Informationen auf der Website über die Site-Suche lohnt, hängt davon ab, wie viel Gewicht man der Kundenfreundlichkeit beimessen will, resp. im Fall der Hochschulwebsite der Studierendenfreundlichkeit, denn die jüngeren Nutzer sind es, die die Suchfunktion am meisten und intuitiv benutzen.

5.5 Googlefication

Als Ausdruck für die „Googlefication“ unserer Gesellschaft (Lebson 2007) prägte Klinkenberg den Ausspruch „There is no need to *know*, because we can *find*“ (Hervorhebung im Original, www.googlification.com). In Anbetracht der Tatsache, dass Finden alles andere als garantiert ist, müsste der Ausspruch präziserweise lauten: „There is no need to *know*, because we can *search*“. Wir können unter dieser Prämisse *das Suchen als neue Form des Zugangs zu Wissen* betrachten, sogar als integraler Bestandteil des Wissenserwerbs. Der Zugang zu Wissen (resp. Informationen) über Suchmaschinen, sowohl im Internet wie auf Websites, ist heute die verbreitetste Form der aktiven Informationserschliessung geworden, d.h. der erste Schritt der Wissensaneignung. Wie wir gesehen haben, ist die Informationserschliessung über Suchmaschinen jedoch alles andere als trivial für die Suchenden. Die Formulierung einer erfolgreichen Suchanfrage, vor allem auf Site-Suchen mit eingeschränkter Menge an Informationsquellen, verlangt vom Suchenden eigentlich ein kognitives Verständnis seines Informationsbedürfnisses und anschliessend eine analytische Transferleistung in ein sprachliches Produkt. Eine erste Grobanalyse des Korpus hat gezeigt, dass diesbezüglich weitergeforscht werden muss, gerade auch im Hinblick auf die Relevanz der Erkenntnisse für das Information Retrieval, nämlich die Umwandlung des Requests in natürlicher Sprache in eine formalisierte Query.

Die Suchbegriffe bilden in dreierlei Hinsicht ein untersuchungswertes linguistisches Korpus (vgl. auch S. 9f), nämlich

- a) Als verbalisierte Deutungsmuster des Informationsbedürfnisses,

- b) Als Ausdruck des Transformationsprozesses in die (vermutete) Logik der Suchmaschine,
- c) Als materielle sprachliche Zeichen mit Kombinationsregeln, die das Retrievalsystem verstehen muss.

Der Aspekt c) wird bereits aktiv erforscht und verschiedene informationslinguistische Verfahren finden Anwendung im Information Retrieval für die Formalisierung der Anfrage als Query. Die Aspekte a) und b) hingegen sind bisher kaum unter dieser Perspektive betrachtet worden. Eine entsprechende Untersuchung müsste auch kognitionswissenschaftliche Konzepte von implizitem Wissen oder Bewusstsein von Nichtwissen oder auch nichtreflektiertem Erfahrungswissen mitverarbeiten. Aus linguistischer Sicht stellt sich die Frage, wie sich bei der Formulierung einer Suchanfrage das Gemeinte und die (semantische) Bedeutung des Geschriebenen zueinander verhalten. Den Suchanfragen fehlt der syntaktische Kontext zur Präzisierung des Gemeinten - komplett im Fall der Ein-Wort-Anfragen, die ja mit rund 70% die grosse Mehrheit der Anfragen ausmachen, aber auch bei den Zwei- und Mehr-Wort-Anfragen. Eine künftige Analyse müsste bei den Mehr-Wort-Anfragen untersuchen, ob durch die Wortkombination eine lexikalische Disambiguierung (Beseitigung der Mehrdeutigkeit) stattfindet. Aufgrund erster Eindrücke wird jedoch vermutet, dass es sich weniger um eine Frage der Disambiguierung als um ein Problem von unpräziser Referenz der isolierten Wörter handelt. Interessanterweise jedoch nicht von Referenz auf die aussersprachliche Realität sondern auf eine „innertextliche“, wenn die gesamte Website als Spezialfall eines Textes betrachtet wird. Dies würde bedeuten, dass die Unterstützung für die Suchenden bei der Formulierung ihrer Anfrage dahin gehen müsste, das Referenzpotenzial der eingetippten Begriffe zu präzisieren resp. zu verkleinern, z.B. indem in irgendeiner Form Metainformationen eingegeben werden können.

Wenn die Informationserschliessung und damit der Zugang zum Wissen heute in immer verstärktem Masse nur noch über die Suchmaschine möglich ist, stellt sich zum Schluss angesichts der Untersuchungsergebnisse also die Frage, ob die Suchfunktion wirklich die Enablerin oder doch eher das Nadelöhr zum Wissen ist ...

Literatur

(Alle Links 10.7.09)

Audrich, S. (2006): The Other Search: Making the Most of Site Search to optimize the Total Customer Experience. Free Digital Marketing Resources, WebSideStory

Braschler, M., Herget, J., Pfister, J., Schäuble, P., Steinbach, M., Stuker, J. (2006): Evaluation der Suchfunktion von Schweizer Unternehmens-Websites. Churer Schriften zur Informationswissenschaft 12 ,
<http://www.computerworld.ch/fileserver/computerworld/files/492.pdf>

Bubenhofer, N. (2006): Einführung in die Korpuslinguistik. Praktische Grundlagen und Werkzeuge. Universität Zürich, Online-Kurs
<http://www.bubenhofer.com/korpuslinguistik/kurs/>

Feldman, S. (2000): Find what I mean, not what I say. Meaning based search tools. In: Online, 24 (3), 49-56

Fries, R. (2007): Informationssuche im Internet. Informationsbedürfnis, Suchmodelle und Suchstrategien. Saarbrücken

Fuhrhop, N. (2008): Das graphematische Wort (im Deutschen): Eine erste Annäherung. In: Zeitschrift für Sprachwissenschaft 27 (2008), 189-228

Griesbaum, J. (2004): Evaluation of three German search engines: Altaviste, Google.de and Lycos. In: Information Research 4

Höchstötter, N. (2007): Suchverhalten im Web – Erhebung, Analyse und Möglichkeiten. In: Information, Wissenschaft & Praxis 58, 3, 135-140, http://www.bui.haw-hamburg.de/pers/ulrike.spree/ws2007_2008/suchverhalten.pdf

Höchstötter, N., Lewandowski, D. (2009): What Users See – Structures in Search engine Result Pages. Preprint accepted for publication in Information Sciences 179, 12, 1796-1812, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/What_users_see_preprint.pdf

Hölscher, C. (1999): Informationssuche im World Wide Web – Messung von Benutzerverhalten. Universität Freiburg, http://gor.de/gor99/tband99/pdfs/a_h/hoelscher.pdf

Hölscher, C., Strube, G.:(2000): Web search behavior of internet experts and newbies. Proceedings of the 9th Int. WWW conference (S.337-346),
<http://www9.org/w9cdrom/81/81.html>.

Kaczmirek, L. (2006): Information und Selektion. Gebrauchstauglichkeit der Ergebnisseiten von Suchmaschinen. Mannheim, Universitätsbibliothek (MADOC) 2006.
<http://bibserv7.bib.uni-mannheim.de/madoc/volltexte/2006/1314/>

- Kaczmirek, L. (2003): Gebrauchtauglichkeit der Ergebnisseiten von Suchmaschinen. In: G. Szwillus, & J. Ziegler (Hrsg.), Mensch & Computer 2003: Interaktion in Bewegung (S. 337-347). Stuttgart, <http://mc.informatik.uni-hamburg.de/konferenzbaende/mc2003/konferenzband/muc2003-33-kaczmirek.pdf>
- Lebson, C. (2007): The „Googlefication“ of Site Search: Adapting to Google’s Influence on User Behavior. Presentes at UPA international 2007
- Lewandowski, D. (2005): Web Information Retrieval. Deutsche Gesellschaft für Informatikwissenschaft und Informationspraxis e.V., Bd 7 (vergriffen), online Version: <http://www.durchdenken.de/lewandowski/web-ir/>
- Lewandowski, D. (2006): Themen und Typen der Suchanfragen an deutsche Web-Suchmaschinen. In: Lehner, F.;Nösekabel, H.; Kleinschmidt, P. (Hrsg): Multikonferenz Wirtschaftsinformatik (MKWI ^06), Universität Passau, Berlin: Gito, S. 33-43
- Lewandowski, D. (2007a): Qualitätsmessung bei Suchmaschinen. System- und nutzerbezogene Evaluationsmasse. In: Informatik Spektrum 30, 3, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/Qualitaetsmessung_bei_Suchmaschine_n_Informatik_Spektrum_2007.pdf
- Lewandowski, D. (2007b): Mit welchen Kennzahlen lässt sich die Qualität von Suchmaschinen messen? In: Machill, M., Beiler, M. (Hrsg.): die Macht der Suchmaschinen / The Power of Search Engines. Köln
- Lewandowski, D., Höchstötter, N. (2008): Websearching: A Quality Measurement Perspective. In: Spink, A.; Zimmer, M. (Hrsg.): Web Searching: Multidisciplinary Perspectives. Berlin, S. 309-340, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/quality_measurement_perspective_2008.pdf
- Lewandowski, D. (2009): The retrieval effectiveness of search engines on navigational queries. Preprint accepted for publication in ASLIB Proceedings 61. http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/ASLIB2009_preprint.pdf
- Nielsen, J. (2001): Search: Visible and Simple. Alertbox May 13, 2001. <http://www.useit.com>
- Notess, G. (2006): Reviews on Google. Auf: Search Engine Showdown. The User’s Guide to Web Searching. <http://searchengineshowdown.com/features/google/review.html>
- Notess, G. (2009): Showdown Analysis, Boolean Searching on Google, Search Engine Showdown. The User’s Guide to Web Searching. <http://www.searchengineshowdown.com/features/google/googleboolean.html>

Quesenbery, W., Jarrett, C., Roddis, I., Stirling, V., Allen, S. (2008a): Search is now normal behavior. What do we do about that? A presentation from: UPA 2008 – The Many Faces of User Experience, June 16-20, 2008, Baltimore, Maryland, USA.
<http://www.wqusability.com/articles/search-is-normal-upa2008.pdf>

Quesenbery, W., Jarrett, C., Roddis, I., Stirling, V., Allen, S. (2008b): Designing for Search: Making Information Easy to Find.
<http://www.wqusability.com/articles/designing-for-search-stc2008.pdf>

Schmidt-Mänz, N., Bomhardt, C. (2005): Wie suchen Onliner im Internet? Science factory 2/2005, <http://www.absatzwirtschaft.de/pdf/sf/Maenz.pdf>

Schmidt-Mänz, N. (2007a). Untersuchung des Suchverhaltens im Web. Interaktion von Internetnutzern mit Suchmaschinen. Studien zum Konsumverhalten, Bd.9. Hamburg

Schmidt-Mänz, N. (2007b): Erkenntnisse aus dem Suchverhalten im Web, „Muster in Suchanfragen“. In: Eberspächer, J., Holtel, S.: Suchen und finden im Internet. Berlin, <http://marketing.wiwi.uni-karlsruhe.de/institut/mitarbeiter/maenz/content/maenz.pdf>

Schmidt-Mänz, N., Koch, M. (2006): A General Classification of (Search) Queries and Terms. Proceedings of International Conference on Information Technologies: Next Generations. Las Vegas, 375-381

Spink, A., Jansen, B.J. (2004): Web Search: Public Searching of the Web, Dordrecht

Spool, J. (2001): People Search Once, Maybe Twice.
http://www.uie.com/articles_search_once/

Stock, W. (2007): Information Retrieval, Informationen suchen und finden. Oldenburg

Stock, W., Stock, M. (2008): Wissensrepräsentation. Informationen auswerten und bereitstellen. München

Véronis, J. (2006): A comparative Study of six search engines. Université de Provence, <http://sites.univ-provence.fr/veronis/pdf/2006-comparative-study.pdf>