

Themen und Typen der Suchanfragen an deutsche Web-Suchmaschinen

Dirk Lewandowski
Institut für Sprache und Information, Abt. Informationswissenschaft
Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1
40225 Düsseldorf
dirk.lewandowski@uni-duesseldorf.de

Abstract: Aufbauend auf Untersuchungen zu den Themen der Websuche und den Anfragetypen (informationsorientiert, navigationsorientiert und transaktionsorientiert) wird eine kombinierte Untersuchung durchgeführt, die neben dem Verhalten der deutschen Suchmaschinen-Nutzer in beiden Bereichen Aussagen über die Bedeutung der einzelnen Anfragetypen in bestimmten Themenfeldern treffen kann.

1 Einleitung

Web-Suchmaschinen stellen ein bedeutendes Feld für die aktuelle Information-Retrieval-Forschung dar [Le05]. Die Kenntnis der Nutzerinteressen ist essentiell für die Betreiber von Informationssystemen allgemein, für die Betreiber von allgemeinen Web-Suchmaschinen im Besonderen. Es ist bekannt, dass im Web nach einer Vielfalt von Themen gesucht wird. Dazu kommen unterschiedliche Anfragetypen: Neben den im klassischen Information Retrieval vorkommenden informationsorientierten finden sich im Web zu nicht unwesentlichen Anteilen auch navigations- und transaktionsorientierte Anfragen.

Ziel der vorliegenden Studie ist es, das Verhalten der deutschen Nutzer sowohl hinsichtlich der gesuchten Themen als auch der Anfragetypen zu untersuchen. Durch eine Kombination der beiden Anfrage-Systematisierungen können Erkenntnisse über die Häufigkeit der einzelnen Anfragetypen in den unterschiedlichen Themenbereichen gewonnen werden.

2 Stand der Forschung

Zum Nutzerverhalten deutscher Suchmaschinen-Nutzer liegt eine umfassende, kombinierte Befragung und Laboruntersuchung [Ma03] vor, in der jedoch keine Erhebung der Themenverteilung vorgenommen wird. In [Hö02] wird das Nutzerverhalten unter anderem auf Basis einer Logfile-Untersuchung der Suchmaschine Fireball betrachtet, allerdings erfolgt auch hier keine thematische Zuordnung. Beide Untersuchungen legen den Schwerpunkt eher auf die Recherchestrategien der Nutzer als auf ihre thematischen Interessen.

Die nachfolgende Untersuchung folgt in ihrem Design den zwei im Folgenden vorgestellten Studien zu Anfragethemen und Anfragetypen. Diese Untersuchungen werden kombiniert, um zu spezifischeren Aussagen über die Bedürfnisse der deutschen Web-Nutzer zu gelangen.

2.1 Themen der Web-Suche

Die Untersuchung der Themen der Websuche basiert auf den Untersuchungen von Spink und Jansen [Sp01] und [SJ04].

Die Themenuntersuchung ist Logfile-basiert, wobei die Themen nur manuell bestimmt werden können. Logfiles bieten den Vorteil, dass aus ihnen tatsächliche Suchanfragen ermittelt werden können, ohne dass der Nutzer in seinem Suchverhalten durch eine Untersuchungssituation beeinflusst wird. Allerdings lassen sich damit auch nur die Anfragen ermitteln; direkte Rückschlüsse auf das jeweilige Informationsbedürfnis sind nicht möglich.

Bei einer solchen qualitativen Untersuchung besteht die Gefahr, dass die Zuordnung der Anfragen zu den Themen von persönlichen Annahmen beeinflusst wird. Dem versuchen die Autoren zu begegnen, indem jede Zuordnung von zwei Personen vorgenommen wird und die Fälle, in denen sich die Urteile unterscheiden, in einer anschließenden Diskussion geklärt werden.

Die Klassifikation der Themen wird in [SJ01] vorgenommen. Es werden die folgenden 11 Klassen gebildet:

- People, places and things (PPT; Personen, Orte und Dinge)
- Computers or Internet (CI; Computer und Internet)
- Commerce, travel, employment, or economy (CTEE; Handel, Reise, Arbeit und Wirtschaft)
- Entertainment or recreation (ER; Unterhaltung und Freizeit)
- Health or sciences (HS; Gesundheit und (Natur-)Wissenschaft)
- Sex or pornography (SP; Sex und Pornographie)
- Government (GOV; Regierung und Verwaltung)
- Education or humanities (EH; Bildung und Geisteswissenschaften)
- Society, culture, ethnicity, or religion (SCER; Gesellschaft, Kultur, Ethnizität und Religion)
- Performing or fine arts (PFA; Kunst)
- Unknown or other (UNK; nicht bestimmbar und sonstige)

Die Klassifikation wird über mehrere Jahre hinweg auf Daten aus Logfiles verschiedener Suchmaschinen angewendet [SJ04]; in der Regel werden etwa 2.500 Suchanfragen ausgewertet.

Im Lauf des Untersuchungszeitraums (1997-2002) stellen die Autoren eine Verschiebung weg von Entertainment-Inhalten hin zu eher wirtschaftsorientierten Anfragen fest. Die meistbesetzten Kategorien in den neueren Untersuchungen sind *People, places or things* (2002: zwischen 41 und 49 Prozent), *Computers or Internet* (2002: zwischen 12 und 16 Prozent) und *Commerce, travel, employment, or economy* (2002: etwa 12 Prozent).

2.2 Typen von Suchanfragen

Eine Unterteilung von Web-Suchanfragen nach Typen nimmt [Br02] vor. Es werden drei Typen identifiziert: *Navigational* (navigationsorientiert), *informational* (informationsorientiert) und *transactional* (transaktionsorientiert).

Das Ziel *navigationsorientierter Anfragen* ist es, eine bestimmte Seite (wieder) zu finden, die dem Benutzer bereits bekannt ist oder von der er zumindest die Vorstellung hat, dass diese existiert. Beispiele für solche Anfragen sind die Suche nach Homepages von Unternehmen („Daimler Chrysler“) oder nach Personen („Bill Gates“). Solche Anfragen haben in der Regel *ein* richtiges Ergebnis; das Informationsbedürfnis ist befriedigt, sobald die gewünschte Seite gefunden wurde. Bei *informationsorientierten Anfragen* ist das Informationsbedürfnis dagegen oft nicht durch ein einziges Dokument zu befriedigen. Vielmehr möchte sich der Nutzer zu einem Thema informieren und liest zu diesem Zweck mehrere Dokumente. Nach Broder zielen informationsorientierte Anfragen auf jeden Fall auf statische Dokumente, d.h. es ist nach dem Aufruf des Dokuments keine weitere Interaktion auf der Website nötig, um an die gewünschten Informationen zu gelangen.

Die dritte Gruppe bilden die *transaktionsorientierten Anfragen*. Der Zweck dieser ist das Auffinden einer Website, auf der dann eine Transaktion stattfinden soll. Transaktionen können beispielsweise der Kauf eines Produkts, der Download einer Datei oder die Recherche in einer Datenbank sein.

Broder untersucht den Anteil der verschiedenen Anfragetypen sowohl anhand einer Nutzerbefragung als auch einer Logfile-Auswertung von 400 Suchanfragen. Beide Untersuchungen beziehen sich auf Daten der Suchmaschine AltaVista.

Aus der Kombination der Ergebnisse beider Erhebungen ergibt sich ein Anteil navigationsorientierter Anfragen von 20-24,5 Prozent, die informationsorientierten Anfragen liegen zwischen 39 und 48 Prozent, die transaktionsorientierten zwischen 22 und 36 Prozent

Die Dreiteilung der Anfragen nach Typen ist als wegweisend anzusehen; spätere Arbeiten versuchen unter anderem, automatisch die hinter einer Anfrage stehende Nutzerintention zu ermitteln. Kang und Kim [KK03] stellen ein Verfahren vor, das Anfragen auf der Basis eines Abgleichs mit zwei Dokumentkolektionen (eine mit typischen Antworten auf navigationsorientierte Anfragen, eine auf informationsorientierte Anfragen gerichtete) entsprechend einer Intention zuordnet, um schließlich auf der Basis einer jeweils unterschiedlichen Relevanzbewertung zu präziseren Retrievalergebnissen zu gelangen.

Eine Erweiterung bzw. Präzisierung von Broders Klassifikation findet sich in [RL04]. Zu jeder Klasse werden einige Unterklassen gebildet; die Klasse „transactional“ wird präzisiert und als eher „ressourcenorientiert“ angesehen. Der Prozess der Klassifikation (der in [Br02] nicht näher beschrieben wird) wird durch ein Tool, welches neben der eigentlichen Suchanfrage die weiteren Interaktionen des jeweiligen Nutzers wie die Trefferauswahl für den Klassierenden bereitstellt, präzisiert.

Im Großen und Ganzen lassen sich die Ergebnisse auf der obersten Hierarchieebene vergleichen, wenn auch einige Verschiebungen auf ein offensichtlich etwas anderes Verständnis der navigations- bzw. transaktionsorientierten Anfragen zurückzuführen sind. Nach [RL04] liegen die informationsorientierten Anfragen bei 61 bis 63 Prozent, die transaktionsorientierten Anfragen (bzw. die *ressources*) bei 21 bis 27 Prozent und die navigationsorientierten Anfragen bei nur 11 bis 15 Prozent.¹

3 Ziele der Untersuchung

Der Zweck der vorliegenden Untersuchung ist es zum einen, die oben beschriebenen Untersuchungsmethoden auf Anfragen an deutsche Suchmaschinen anzuwenden, um das typische Nutzerinteresse der deutschen Suchmaschinennutzer herauszufinden und mit dem der in den beschriebenen Studien untersuchten Verhalten der zumeist US-amerikanischen Nutzer zu vergleichen.

Zum anderen soll durch die Kombination der beiden Untersuchungsmethoden herausgefunden werden, ob sich in bestimmten Themenfeldern bestimmte Anfragetypen häufen. Letztlich stellt sich die Frage, ob Suchmaschinen Anfragen der jeweiligen Themenfelder adäquat behandeln.

In Anbetracht des Ziels der Kombination der Anfragethemen und -typen wird auf eine Erweiterung bzw. Präzisierung der Anfragetypen wie oben beschrieben verzichtet. Für eine solche feingliedrige Darstellung wäre eine sehr große Anfragemenge zu untersuchen, was im Rahmen dieser Studie nicht geleistet werden kann.

4 Methode

Für die Untersuchung wurden die sich explizit auf den deutschen Markt richtenden Suchmaschinen Fireball <www.fireball.de> und Seekport <www.seekport.de>, dazu die Metasuchmaschine Metager <www.metager.de>, ausgewählt. Alle untersuchten Anfragen stammen vom 18. Oktober 2005.

Von Metager konnten die kompletten Logfiles bezogen werden, so dass hier eine zufällige Auswahl der Suchanfragen leicht gewährleistet werden konnte. Da ein Zugriff auf die Logfiles bei Fireball und Seekport nicht möglich war, musste hier auf die bei beiden Suchmaschinen vorhandene Live-Suche² zurückgegriffen werden. Dabei handelt es sich um eine Anzeige der aktuell an die Suchmaschine gestellten Suchanfragen. Diese Seiten wurden über den Tag verteilt immer wieder aufgerufen und die angezeigten Suchanfragen kopiert. Damit konnte wenigstens näherungsweise eine repräsentative Auswahl der Suchanfragen über den Tag verteilt ermittelt werden.

¹ Die Unterschiede ergeben sich aus den insgesamt drei von Rose u. Levinson durchgeführten Untersuchungen.

² <http://www.fireball.de/livesuche/> bzw. <http://www.seekport.de/q?liveseek>

Andere Suchmaschinen – allen voran die „großen drei“: Google, Yahoo und MSN – bieten keine Möglichkeit der Einsicht in momentan gestellte Anfragen. Sie konnten deshalb in die Untersuchung nicht eingebunden werden.

Schon während der Erfassung der Daten zeigte sich, dass Seekport offensichtlich Anfragen, die sich auf Pornographie beziehen, filtert und in der Live-Suche nicht anzeigt. Daher ist für diesen Bereich mit einer gewissen Verzerrung der Ergebnisse zu rechnen.

Von jeder Suchmaschine wurden 500 Anfragen ausgewählt, so dass insgesamt 1.500 Anfragen ausgewertet wurden. Gegenüber der Untersuchung von Broder mit 400 manuell klassifizierten Suchanfragen wird damit eine deutlich größere Anfragemenge verwendet; gegenüber der Themenklassifikation in [SJ04] von jeweils etwa 2.500 untersuchten Anfragen ist die ausgewertete Menge allerdings deutlich geringer. Die verwendete Datenbasis ist jedoch ausreichend, um zu validen Aussagen zu gelangen.

Jede Suchanfrage wurde von zwei Juroren jeweils einem Thema und einem Anfragetyp zugeordnet. Unstimmigkeiten zwischen den Juroren wurden diskutiert und in einer gemeinsamen Entscheidung beigelegt. Damit sollte der Problematik der subjektiven Bewertung der Suchanfragen begegnet werden. Generell bleibt die Problematik bestehen, aus den meist recht kurzen Anfragen aus den Logfiles auf das dahinterstehende Informationsbedürfnis des Nutzers zu schließen.

5 Ergebnisse

5.1 Anfragetypen

Die größte Gruppe bilden die informationsorientierten Anfragen, die über alle Suchmaschinen verteilt bei 45 Prozent aller Anfragen liegen, wobei die Werte je nach Suchmaschine zwischen 42 und 47 Prozent liegen. In einem ähnlichen Rahmen bewegen sich die Werte der navigationsorientierten Anfragen, die einen Durchschnitt von 40 Prozent erreichen. Größere Abweichungen finden sich bei den transaktionsorientierten Anfragen. Diese liegen zwischen 11 und 18 Prozent mit einem Durchschnitt von 15 Prozent. Die genaue Verteilung der Daten nach Suchmaschinen findet sich in Tabelle 1.

Tabelle 1. Verteilung der Anfragen nach Typen in den untersuchten Suchmaschinen

Suchmaschine	Informationsorientiert	navigationsorientiert	transaktionsorientiert
Fireball	47%	35%	18%
Metager	42%	43%	15%
Seekport	47%	42%	11%
Durchschnitt	45%	40%	15%

Es zeigt sich also, dass alle drei Anfragetypen einen nennenswerten Anteil am Gesamt der Anfragen auf sich ziehen können. Für Suchmaschinen ist es also wichtig, alle drei Anfragetypen beantworten zu können. Vergleicht man die Daten mit der Untersuchung von Broder [Br02], so zeigen sich deutliche Abweichungen: Während die informationsorientierten Anfragen in dem von Broder ermittelten Rahmen liegen, zeigt die Auswertung einen höheren Anteil navigationsorientierter Anfragen, während die transaktionsorientierten Anfragen einen deutlich geringeren Anteil ausmachen. Die Gründe für diese Unterschiede lassen sich nicht klar bestimmen: Die Veränderungen könnten aus dem zeitlichen Abstand zwischen den Untersuchungen resultieren, aber auch aus dem spezifischen Nutzerverhalten deutscher Nutzer bzw. der von Broder untersuchten Nutzer der Suchmaschine AltaVista. Es wäre auch plausibel, den geringeren Anteil transaktionsorientierter Anfragen auf die zunehmende Verwendung von speziellen Shopping-Suchmaschinen zurückzuführen, welche speziell auf diesen Anfragetyp ausgerichtet sind.

5.2 Themen

Abbildung 1 zeigt die Verteilung der Themen für alle Suchmaschinen der aktuellen Untersuchung im Vergleich zu den zwei aktuellsten in [SJ04] beschriebenen Untersuchungen.

Von den deutschen Suchmaschinen-Nutzern 2005 werden am stärksten Themen aus der Wirtschafts- und Arbeitswelt nachgefragt (29 Prozent aller Anfragen), gefolgt von der Suche nach Personen und Orten (12,8 Prozent). Alle anderen Themenklasse liegen deutlich unter 10 Prozent; zwischen sieben und acht Prozent jeweils Computer und Internet, Unterhaltung und Freizeit sowie Gesundheit und Wissenschaft. Die restlichen Themen liegen zwischen einem und etwas mehr als vier Prozent. Die Themen sind zwar unterschiedlich stark, jedoch breit verteilt. Auch auf die am schwächsten besetzte Kategorie (Kunst) entfällt immerhin noch etwa jede hundertste Anfrage.

Die Verteilung unterscheidet sich deutlich von den Vergleichsuntersuchungen, in denen jeweils mit großem Abstand der Themenbereich Personen, Orte und Dinge führt (41 bzw. 49Prozent), mit großem Abstand gefolgt von Computer und Internet sowie Wirtschaft und Arbeit. Wiederum lassen sich die Gründe für die Unterschiede nicht verlässlich bestimmen: Sowohl zeitliche Unterschiede als auch Unterschiede im Nutzerverhalten wären plausibel. Deutlich ist auf jeden Fall der hohe Anteil der wirtschaftsorientierten Anfragen der deutschen Nutzer; hier wäre zu beobachten, ob sich durch direktere Suchmöglichkeiten wie etwa durch Shopping-Suchmaschinen Veränderungen ergeben.

Der Anteil der Anfragen nach Sex und Pornographie liegt bei durchschnittlich 4,5 Prozent, wobei berücksichtigt werden muss, dass solche Suchanfragen in der Livesuche bei Seekport offensichtlich gefiltert werden. Bei Metager liegt der Anteil dieser Anfragen bei 3,2 Prozent, bei Fireball deutlich höher bei 10,2 Prozent.

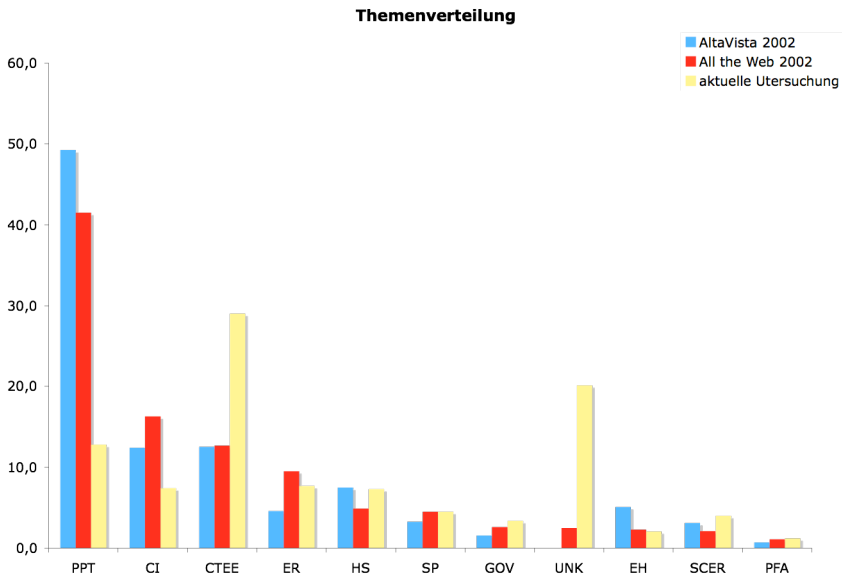


Abbildung 1: Verteilung der Anfragen nach Themen (alle Suchmaschinen)

5.3 Anfragetypen innerhalb der Themenblöcke

Betrachtet man die informationsorientierten Anfragen (Abb. 2), so zeigen sich insbesondere zwei Themen, die weit über dem Durchschnitt aller Anfragen liegen, nämlich Anfragen aus dem Bereich Gesundheit und Wissenschaft sowie Gesellschaft, Kultur, Ethnizität und Religion. Am schwächsten besetzt sind die Themen Personen und Orte sowie Computer und Internet.

Die Suche nach Personen erfolgt deutlich öfter navigationsorientiert (s. Abb. 3). Dies ist aber zumindest zum Teil auch auf die verwendete Methodik zurückzuführen, die Anfragen nach Personen, die keine Prominenten sind, als Homepagesuche klassifiziert und damit den navigationsorientierten Anfragen zurechnet. Weitere Themen, in denen mehr als 40 Prozent der Anfragen navigationsorientiert sind, sind Wirtschaft und Arbeit, Unterhaltung und Freizeit, Regierung und Verwaltung sowie Kunst.

Am deutlichsten sind die Unterschiede zwischen den Themen wohl beim Anfragetyp „transaktionsorientiert“ (s. Abb. 4). Während drei Kategorien keinen nennenswerten Anteil an transaktionsorientierten Anfragen haben, liegt der Anteil in drei Kategorien über 25 Prozent. Dies sind Computer und Internet (34,8 Prozent), Wirtschaft und Arbeit (26,3 Prozent) sowie Sex und Pornographie (25 Prozent).

Anteil informationsorientierter Anfragen nach Themen

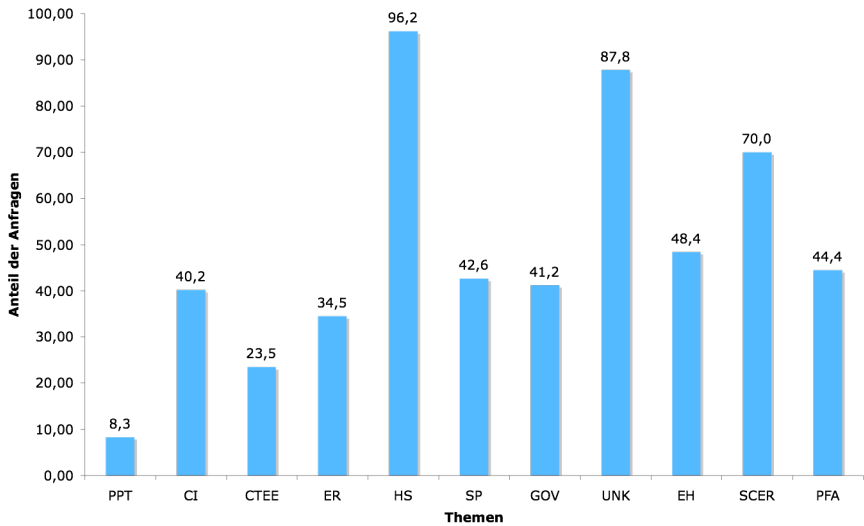


Abbildung 2: Anteil informationsorientierter Anfragen nach Themen

Anteil navigationsorientierter Anfragen nach Themen

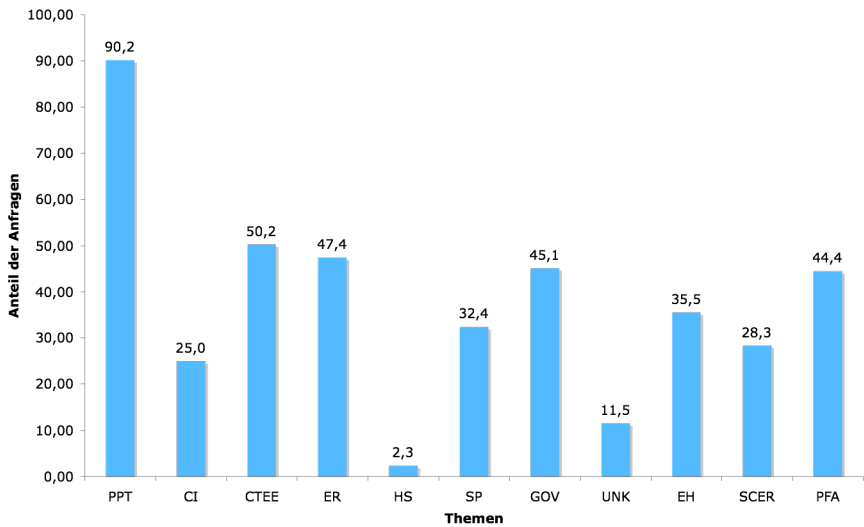


Abbildung 3: Anteil navigationsorientierter Anfragen nach Themen

Anteil transaktionsorientierter Anfragen nach Themen

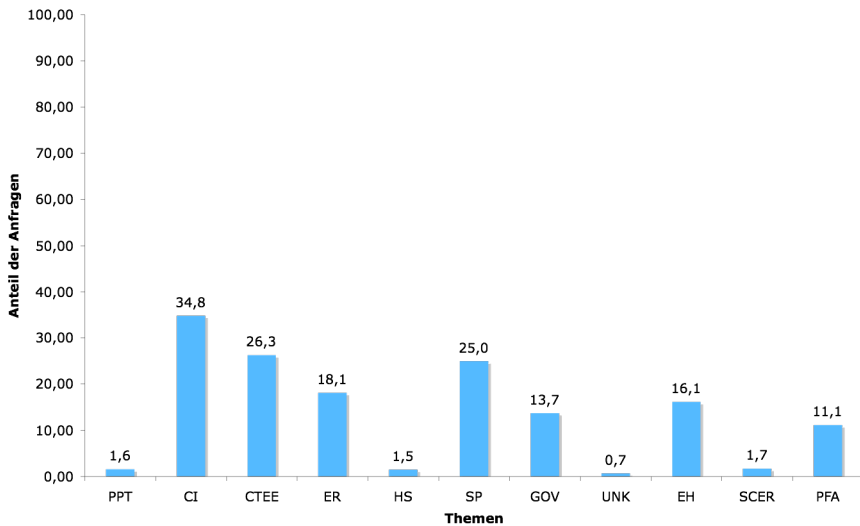


Abbildung 4: Anteil transaktionsorientierter Anfragen nach Themen

Im Bereich Computer werden hier vor allem Anfragen nach dem Download von Programmen gestellt, bei Sex und Pornographie sind es vor allem Anfragen nach dem Download von Filmen. Interessant ist, dass auch im Bereich Unterhaltung und Freizeit etwa 18 Prozent der Anfragen transaktionsorientiert sind, d.h. dass ein Interesse der Nutzer besteht, vor allem Filme oder Musik direkt herunterzuladen.

6 Diskussion

Konsequenzen für den Aufbau von Suchmaschinen ergeben sich einerseits aus der Verteilung der Anfragetypen und andererseits aus deren Verteilung innerhalb der Themenblöcke. Suchmaschinen sollten alle Anfragetypen berücksichtigen, da jeder Typ einen signifikanten Anteil am Gesamt der Suchanfragen stellt. Linktopologische Rankingverfahren, wie sie bei den heutigen Suchmaschinen als wesentlicher Rankingfaktor eingesetzt werden, entfalten ihre Stärken vor allem bei navigationsorientierten Anfragen. Auch in der Evaluierung von Suchmaschinen werden die unterschiedlichen Anfragetypen bisher nur unzureichend berücksichtigt; Retrievaltests konzentrieren sich meist auf die informationsorientierten Anfragen.

In Bezug auf die Verteilung innerhalb der Themenblöcke zeigt sich, dass es für Suchmaschinen lohnend sein könnte, bei bestimmten Themen Unterstützung für einzelne Anfragetypen einzuführen. Denkbar ist an den automatisierten Hinweis auf Spezialbestände für den Bereich Computer (etwa auf Download-Verzeichnisse für Software) oder eine spezielle Audio- oder Videosuche für den Themenbereich Unterhaltung.

Ob die von den Suchmaschinen angezeigten Ergebnisse den Nutzererwartungen entsprechen, konnte im Rahmen dieser Untersuchung nicht geklärt werden, ist aber sicher ein lohnenswertes Thema für zukünftige Studien. So stellt sich etwa die Frage, ob die in großer Mehrheit informationsorientierten Anfragen etwa aus dem Bereich Gesundheit und Wissenschaft sowie Gesellschaft von den Suchmaschinen auch entsprechend beantwortet werden oder ob die Ergebnisse hier eher auf Transaktionen ausgerichtet sind.

Die vorgelegte Untersuchung kann nur einen Anfang machen, das Nutzerverhalten bezüglich der Themen und Anfragetypen zu klassifizieren. Die vorhandenen Systematiken wurden verwendet, um eine Vergleichbarkeit zu gewährleisten. Für zukünftige Untersuchungen wäre vor allem eine genauere thematische Unterteilung sinnvoll. In Ergänzung sollten Nutzer zu ihren Anfrageintentionen direkt befragt werden, um durch einen Vergleich die Zuverlässigkeit der teils schwierigen Zuordnung der Anfragen zu verbessern.

Literaturverzeichnis

- [Br02] Broder, A. (2002). A taxonomy of web search. SIGIR Forum, 36(2), 3-10.
- [Hö02] Hölscher, C. (2002). Die Rolle des Wissens im Internet. Gezielt suchen und kompetent auswählen. Stuttgart: Klett-Cotta.
- [KK03] Kang, I.-H., & Kim, G. (2003). In Query Type Classification for Web Document Retrieval (pp. 64-71). Paper presented at the SIGIR'03, Toronto, Canada. ACM Press.
- [Le05] Lewandowski, D. (2005). Web Information Retrieval: Technologien zur Informationssuche im Internet. Frankfurt am Main: DGI.
- [Ma03] Machill, M., Neuberger, C., Schweiger, W., & Wirth, W. (2003). Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen. In M. Machill & C. Welp (Eds.), Wegweiser im Netz. Gütersloh: Bertelsmann Stiftung.
- [RL04] Rose, D.E., & Levinson, D. (2004). Understanding user goals in Web search. Thirteenth International World Wide Web Conference Proceedings, WWW2004, 13-19.
- [SJ04] Spink, A., & Jansen, B.J. (2004). Web Search: Public Searching of the Web (Vol. 6). Dordrecht: Kluwer.
- [Sp01] Spink, A., Wolfram, D., Jansen, B.J., & Saracevic, T. (2001). Searching the Web: The Public and Their Queries. Journal of the American Society for Information Science and Technology, 53(3), 226-234.