



Hochschule für Angewandte Wissenschaften Hamburg  
*Hamburg University of Applied Sciences*

## **Diplomarbeit**

Zur Struktur der Trefferbeschreibungen in Web-Suchmaschinen und deren Beurteilung durch den Nutzer

vorgelegt von

***Lukas Tittmann***

*am 15. August 2008*

Erstbetreuer: Prof. Dr. Lewandowski

Zweitbetreuer: Prof. Dr. Spree

## Abstract

Gegenstand der Arbeit sind Trefferbeschreibungen von Web-Suchmaschinen. Diese werden unter der Berücksichtigung der verschiedenen Quellen, aus denen sie zusammengestellt oder extrahiert werden können, im Hinblick auf verschiedene Merkmale wie Länge, sprachliche Beschaffenheit und Aussagekraft für den Nutzer betrachtet.

Die verschiedenen Suchmaschinen nutzen unterschiedliche Quellen für ihre Trefferbeschreibungen, z.B. Ausschnitte aus dem Inhalt der einzelnen Dokumente (*Keyword in Context* = KWIC), die Metadaten der Dokumente oder Beschreibungen aus externen Verzeichnissen. In einer vergleichenden Studie werden mehrere Suchanfragen diese Quellen sowie weitere Eigenschaften wie Länge oder sprachliche Beschaffenheit der Trefferbeschreibungen erhoben. Mittels schriftlicher Nutzerbefragung wird zusätzlich die Frage nach der Beschreibungsqualität untersucht, d.h. die Frage, welche Trefferbeschreibungen von den Nutzern als besonders aussagekräftig empfunden werden.

### *Kernergebnisse:*

Derzeit werden KWIC-Beschreibungen am häufigsten von Suchmaschinen genutzt, allerdings erhalten die anderen Quellen bessere Bewertungen von den Nutzern. Als Schlüsselfaktor für die Aussagekraft einer Beschreibung wird die sprachliche Qualität identifiziert.

Anhand der Erkenntnisse der Studie werden ein Konzept zur Erstellung von Beschreibungen entworfen sowie weitere Verbesserungsmöglichkeiten in Bezug die Aussagekraft für den Nutzer aufgezeigt.

### Schlagwörter:

Web-Suchmaschinen, Trefferbeschreibungen, Beschreibungen, Suchmaschinen, Treffer, Vergleichsstudie, Studie, Beschreibungsqualität.

## Inhalt

|   |    |
|---|----|
| Einleitung.....   | 1  |
| 1 Begriffsdefinition/Abgrenzung.....  | 3  |
| 1.1 Der Suchprozess bei Websuchmaschinen.....                                       | 3  |
| 1.2 Trefferseiten (SERPs) .....   | 4  |
| 1.3 Trefferbeschreibungen auf SERPs .....   | 5  |
| 1.4 Typen/Quellen der Trefferbeschreibungen .....                                   | 7  |
| Keywords in Kontext.....  | 8  |
| Metadaten aus dem Quelltext .....   | 8  |
| Verzeichnisse .....   | 9  |
| 2 Studien zu Trefferbeschreibungen.....   | 9  |
| 2.1 Studie Lewandowski 2008 .....   | 12 |
| 2.2 Crystal und Greenberg 2006 .....  | 13 |
| 2.3 Zusammenfassende Würdigung der Studien .....                                    | 14 |
| 3 Untersuchung der Trefferbeschreibungen und deren Bewertung.....                   | 15 |
| 3.1 Fragestellungen .....   | 16 |
| 3.2 Methodik .....  | 17 |
| Erhebung der Daten .....  | 17 |
| Zu erfassende Merkmale .....  | 18 |
| Fragebogen .....  | 19 |
| 3.3 Durchführung Analyse der Trefferbeschreibungen .....                            | 21 |
| 3.4 Durchführung der Bewertung mit Fragebögen .....                                 | 23 |
| 3.5 Ergebnisse .....  | 24 |
| 3.5.1 Analyse der Trefferbeschreibungen.....  | 24 |
| 3.5.2 Untersuchung der Wikipedia-Treffer .....                                      | 32 |
| 3.5.3 Zusammenfassung der Strukturanalyse der<br>Trefferbeschreibungen.....         | 34 |
| 3.5.4 Bewertung der Treffer mit Fragebögen .....                                    | 36 |
| 3.5.5 Zusammenfassung der Befragung zur Bewertung der<br>Trefferbeschreibungen..... | 44 |
| 3.6 Diskussion der Untersuchungen .....   | 46 |

|   |  |    |
|---|--|----|
| 4 | Entwicklung einer idealen Trefferbeschreibung.....             | 48 |
| 5 | Fazit .....  | 54 |
| 6 | Literatur und Quellenverzeichnis .....                         | 56 |
|   | Anhang.....  | 61 |
|   | A1. Suchanfragen.....  | 61 |
|   | A2. Fragebogen.....  | 63 |
|   | A3. Zusammensetzung der Bewertungssets .....                   | 67 |
|   | A4. Paarungen von Beschreibungstypen im 1. Und 2. Element..... | 70 |
|   | A5. Länge der Trefferbeschreibungen (in Wörter).....           | 71 |
|   | A6. Länge der Trefferbeschreibungen (in Zeichen) .....         | 72 |
|   | A7. Anzahl der Zeilen der Beschreibungen.....                  | 73 |
|   | A8. Anteile Sprachtypen je Suchmaschine .....                  | 74 |
|   | A9. Anteile Sprachtypen je Beschreibungsquelle.....            | 75 |
|   | A10. Anteile an Wikipedia Treffer .....                        | 75 |
|   | A11. Bewertungen und Länge der Beschreibungen.....             | 75 |

In dieser Arbeit werden folgende Abkürzungen verwendet:

KWIC- Keywords in Context  
ODP- Open Directory Project  
SA- Suchanfrage  
SERP- Search Engine Result Page  
SM- Suchmaschine  
TB- Trefferbeschreibung

## **Einleitung**

Suchmaschinen sind eine der Anwendungen, die im Internet am häufigsten genutzt werden. Die jüngste ARD/ZDF Onlinestudie stellte fest, dass Suchmaschinen sogar vor dem „Klassiker“ Emails schreiben und empfangen am häufigsten genutzt werden (van Eimeren u.a. 2008, S. 337). In der selben Studie wird festgestellt, dass das Internet überwiegend dazu genutzt wird, um Informationen zu erhalten. Betrachtet man die Zahlen, wie viele Treffer für eine Suchanfrage gefunden werden, so wird deutlich, dass nur ein sehr kleiner Anteil der ermittelten Treffer gesichtet werden kann. Selbst wenn man sich nur auf die ersten zwei Trefferseiten beschränkt, sind es noch 20 Treffer, mit denen man sich auseinandersetzen muss. Daraus folgt, dass eine gewisse Vorauswahl, welche Treffer für die weitere Auswertung in Frage kommen, getroffen werden muss. In dieser Phase spielen die Trefferbeschreibungen eine entscheidende Rolle. So ist der Beschreibungstext neben dem Titel des Treffers das Element einer Trefferseite, das am meisten beachtet wird, (Machill u.a. 2003, S. 171). Dies kann jedoch nur dann gelingen, wenn die Trefferdokumente repräsentativ und aussagekräftig beschrieben werden, da sonst keine Rückschlüsse auf den Inhalt und die Qualität eines Dokuments möglich sind. Umso erstaunlicher ist festzustellen, dass die Trefferbeschreibungen innerhalb der Suchmaschinenforschung noch kaum gewürdigt wurden. Zwar setzen sich einige wenige Studien mit der Retrievaleffektivität der Trefferbeschreibungen auseinander, aber es liegen keine Kenntnisse vor, welche Quellen der Trefferbeschreibungen von den Suchmaschinen genutzt werden und wie die Trefferbeschreibungen strukturell beschaffen sind. Mit der vorliegenden Ausarbeitung soll ein erster Schritt getan werden, dieses Defizit aufzuholen.

Diese Arbeit ist in vier Bereiche untergliedert. Im ersten Abschnitt soll auf die verschiedenen Arten von Trefferbeschreibungen eingegangen werden. Darauf folgt eine kurze Betrachtung bisheriger Studien, die sich mit Trefferbeschreibungen auseinandersetzen. Den Kern dieser Arbeit bildet eine Untersuchung der Trefferbeschreibungen, die sich in zwei Abschnitte auf-

teilt: Im ersten werden die Struktur der Trefferbeschreibungen und die derzeitige Verwendung in den Suchmaschinen analysiert. Der zweite Abschnitt setzt sich mit der Bewertung der Trefferbeschreibungen durch den Nutzer auseinander. Für die Bewertungen der Trefferbeschreibungen wurden Studenten der Departments Information sowie Wirtschaft befragt<sup>1</sup>.

Abgerundet wird die Arbeit durch die Entwicklung eines Schemas zur Generierung der Trefferbeschreibungen das auf den Erkenntnissen der Untersuchungen basiert, sowie weitere Verbesserungsmöglichkeiten aufzeigt.

---

<sup>1</sup> An dieser Stelle soll Herrn Prof. Dr. Lewandowski und Frau Prof. Kullmann für ihre Bereitschaft, ihre Seminare für die Befragung zur Verfügung gestellt haben, gedankt werden.

# **1 Begriffsdefinition/Abgrenzung**

In den folgenden Abschnitten sollen zum einen die grundlegenden Begrifflichkeiten erläutert und zum anderen der Bezugsrahmen der Themenstellung abgegrenzt werden.

## **1.1 Der Suchprozess bei Websuchmaschinen**

Nach Marchionini lässt sich innerhalb des Suchprozesses die Interaktion mit einer Web-Suchmaschine in mehrere Phasen unterteilen, darunter: a) Wahl der Suchmaschine b) Formulierung der Suchanfrage c) Auslösen der Suche und d) Evaluation der Suchergebnisse (Marchionini 1995, S. 50 ff.). Der Fokus dieser Arbeit beschränkt sich auf die letztgenannte Phase, die Evaluation der Suchergebnisse. Tang und Solomon bzw. Rieh betrachten diese Phase noch differenzierter, in dem sie diese nach zwei Aspekten aufteilen: der Evaluation der Trefferbeschreibungen auf der Trefferseite und der Evaluation der Dokumente selbst (Tang u.a. 2001, S. 676; Rieh 2002, S. 146 f.). Es wird also einerseits zwischen der Evaluation der Treffer auf den Suchergebnisseiten (die Auszüge oder Beschreibungen der Dokumente enthalten) und andererseits den Dokumenten unterschieden, zu denen die Treffer verlinkt sind. Der Nutzer trifft anhand der Textauszüge und Beschreibungen der Dokumente eine Vorauswahl und sichtet daraufhin nur jene Dokumente, die der Nutzer anhand der Trefferbeschreibung am ehesten für geeignet hält, dass diese dessen Informationsbedürfnis befriedigen. Daher ist in dieser ersten Phase der Trefferevaluation von entscheidender Bedeutung, wie diese Trefferbeschreibungen beschaffen sind, weil diese einen unmittelbaren Einfluss auf die Frage haben, ob ein Treffer für relevant gehalten wird oder nicht. Ein weiterer Aspekt der Unterscheidung zwischen der Suchergebnisseite und den Dokumenten selbst, ergibt sich aus der Tatsache, dass es sich bei der ersten Stufe noch um einen Bestandteil der Suchmaschine handelt, der auch von der Suchmaschine direkt beeinflusst werden kann, wohingegen sich die Dokumente selbst schon außerhalb des Einflussbereiches der Suchmaschinen befinden. Hölscher betrachtet die Evaluation der Ergebnisse als einen „Dreh- und Angelpunkt des Suchprozesses“, da zu diesem Zeit-

punkt eine (Vor-)Entscheidung getroffen wird, welche Treffer gesichtet werden und welche nicht. Einen weiteren wichtigen Aspekt in der Evaluation der Treffer sieht Hölscher darin, dass die Trefferbeschreibungen wichtige Hinweise für Neuformulierungen der Suchanfrage geben können. So können sich aus dem Inhalt einer Trefferbeschreibung Begriffe ergeben, die der Suchanfrage vergleichbar sind oder auch Anregungen, wie die Suchanfrage eingegrenzt oder erweitert werden kann. (Hölscher 2002, S. 39).

## 1.2 Trefferseiten (SERPs)

Nachdem man eine Suchanfrage abgeschickt hat und diese von der Suchmaschine verarbeitet wurde, erhält man eine Trefferseite, auch SERP (Searchengine result page) genannt. Diese besteht aus mehreren Teilen: Neben der algorithmisch generierten Trefferliste auch bezahlte Treffer (sponsored Links), Werbung, sowie Hinweise auf weitere Kollektionen, wie zum Beispiel Nachrichten, Bilder oder eine lokale Suche. In ihrer Grundstruktur haben die Trefferseiten bei den großen Suchmaschinen einen vergleichbaren Aufbau (Lewandowski 2005, S. 29). Abbildung 1 zeigt einen typischen Aufbau einer Trefferseite. In einem mittleren, etwas nach unten versetzten Bereich werden die eigentlichen Trefferseiten angezeigt [1], die aus einer Abfolge von mehreren einzelnen Treffern bestehen. Auf die Treffer im Einzelnen wird im nachfolgenden Abschnitt detaillierter eingegangen.

Oberhalb und vereinzelt auch unterhalb dieser generierten Treffer werden sponsored Links [2] angezeigt, die meist in einer anderen graphischen Gestaltung vorliegen, als die algorithmisch generierten Treffer. Im rechten Bereich, oder auch über den Treffern, ist oft ein Bereich mit Werbung vorzufinden, der in der Regel mit zur Suchanfrage passenden Anzeigen versehen ist<sup>2</sup>. Wie in der Abbildung gut zu sehen ist, werden nicht alle der 10 regulären Treffer, die die Suchmaschinen standardmäßig pro Trefferseite anzeigen, dargestellt, sondern sind erst durch nach unten scrollen sichtbar. Eine neuere Entwicklung stellt die verstärkte Einbindung weiterer Kollektionen

---

<sup>2</sup> Z.B. das Ad-Sense-Programm von der Firma Google.

in die Ergebnisseite der regulären Web-Suche dar. Lewandowski beschreibt zwei gegensätzliche Ansätze, denen derzeit nachgegangen wird: Einerseits die Einbindung dieser Treffer aus anderen Kollektionen innerhalb der regulären Treffern und andererseits die auf einer gemeinsamen Seite dargestellten, aber optisch deutlich nach Kollektion abgegrenzte Präsentation der Treffer (Lewandowski 2007, S. 84 ff.).

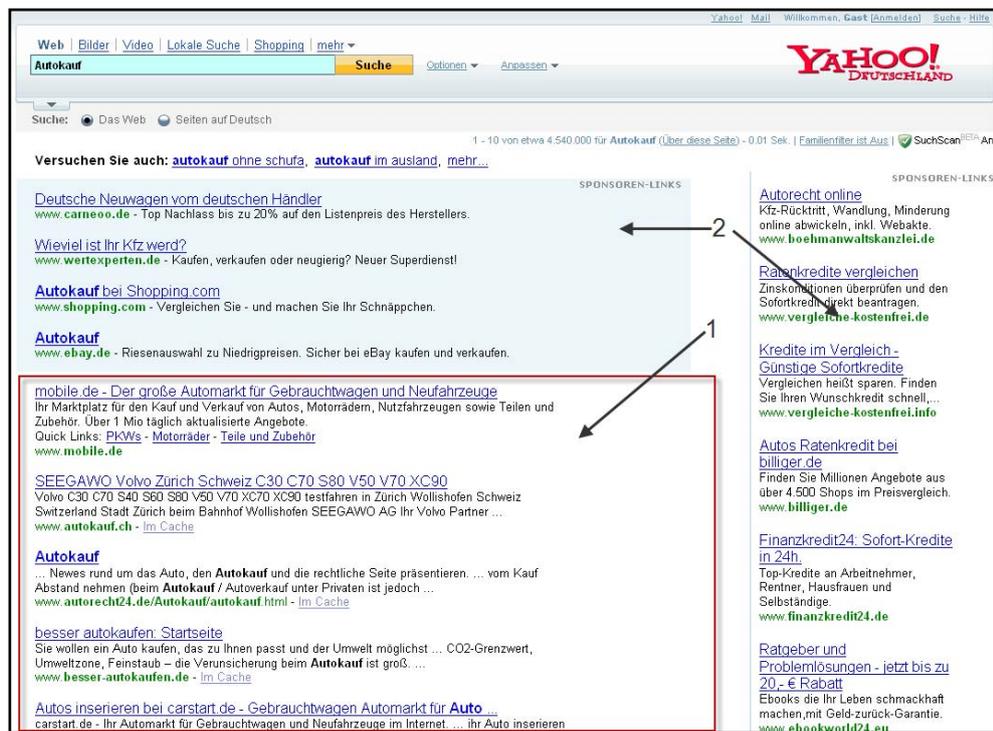


Abbildung 1

### 1.3 Trefferbeschreibungen auf SERPs

Durch die hohe Anzahl an Treffern, die man in der Regel bei den Web-Suchmaschinen erhalten kann, sind die Beschreibungen der einzelnen Treffer für die Auswahl der für den Nutzer relevanten Treffer von grundlegender Bedeutung, weil die Trefferbeschreibungen dem Nutzer wichtige Informationen bieten, um was für ein Dokument es sich handelt, und welchen Inhalt es hat.



**Abbildung 2 : Elemente der Trefferbeschreibung (hier am Beispiel von Google)**

Abbildung 2 zeigt einen typischen Aufbau eines Treffers, der auf eine Webseite, oder auch ein anderes Dokument im Internet verweist. Dieser besteht aus mehreren Elementen: Eingeleitet wird der Treffer mit einem Titel [1], der in der Regel dem Title-Tag der Webseite entnommen wird (Erlhofer 2007b, S. 234). Unterhalb folgt eine Beschreibung des Treffers [2]. Diese kann einen Ausschnitt des Textes, einzelne Wörter der Webseite, den Inhalt der Metadaten, (z. B. den Description-Tag), oder aber auch Angaben zu Webseiten, die aus Verzeichnissen, (wie das Open Directory Project<sup>3</sup> (ODP)), stammen, enthalten. Unterhalb der Beschreibung wird die Web Adresse, auch URL (Uniform Resource Locator) genannt, angezeigt [3]. Diese drei Elemente werden von den meisten Suchmaschinen als Kern angeboten und sind somit als Standard anzusehen (Fauldrath u.a. 2005, S. 26; Aschoff u.a. 2007, S. 146). Je nach Suchmaschine können die Treffer weitere Komponenten enthalten, wie zum Beispiel, eine Angabe zur Größe der Datei, die den Treffer enthält, einen Link zum Cache-Version<sup>4</sup> oder auch einen Hinweis auf ähnliche Treffer, sowie das Datum, das den Zeitpunkt der letzten Erfassung der Webseite durch die Suchmaschine festhält. In der Literatur lassen sich weitere Komponenten finden, die heute jedoch kaum verbreitet sind, darunter fallen zum Beispiel die Trennung zwischen dem Ausschnitt aus dem Text der Webseite, sowie die Beschreibung, die den Verzeichnissen entnommen wird oder auch die Darstellung von Kategorien, die

<sup>3</sup> <http://www.dmoz.de/>

<sup>4</sup> hierbei handelt es sich um eine bei der Suchmaschine hinterlegte Kopie des Dokuments. In den Suchmaschinen wird dieser Service unter verschiedenen Namen angeboten.

einem Treffer zugeordnet werden (Beiler 2005, S. 168). Im Fokus der weiteren Betrachtung eines Treffers steht die Beschreibung, die den Inhalt des Dokuments präsentiert.

#### **1.4 Typen/Quellen der Trefferbeschreibungen**

Um eine Beschreibung zu erhalten, die den Inhalt des Treffers repräsentiert, greifen die Suchmaschinen auf verschiedene Verfahren zurück. Grundsätzlich kann man zwischen Verfahren, bei denen interne Quellen (das heißt, es wird direkt auf das Dokument zurückgegriffen), und externen Quellen (wie Verzeichnisse, die sich außerhalb der Webseite befinden), unterscheiden. Drori ordnet beiden Varianten verschiedene Methoden zu: Zu den Verfahren bei denen auf interne Quellen zurückgegriffen wird, zählen die Extraktion a) signifikanter Sätze b) signifikanter (einzelner) Wörter c) Informationen aus Metadaten des Dokuments. Den externen Verfahren werden d) die Klassifikation des Dokumentes e) die auf ein Dokument zitierenden Dokumente f) Informationen aus der Datenbank, in der das Dokument vorliegt, zugeordnet (Drori 2000, S. 73 ff.). Lewandowski beschreibt die verwendeten Methoden etwas konkreter und nennt das Keywords in Context Verfahren (KWIC), die Beschreibung aus dem Meta-Description Tag, sowie die Beschreibungen aus externen Quellen, wie zum Beispiel Verzeichnisse (Lewandowski 2005, S. 224 f.).

Das KWIC Verfahren ist damit eine Bündelung der signifikanten Sätze und Wörter. Der Unterschied besteht insbesondere darin, dass beim KWIC-Ausschnitt Suchbegriffe entgalten sind. Die externen Quellen, wie zum Beispiel ein Verzeichnis, können nach Droris Einordnung als ein Spezialfall der zitierenden Dokumente gesehen werden, da die Verzeichnisse mit einem Link auf die entsprechenden Dokumente verweisen. Im Folgenden sollen drei Methoden genauer betrachtet werden: Die Textextraktion, die Nutzung der Metadaten eines Dokuments sowie die Verzeichnisse.

## **Keywords in Kontext**

Hierbei handelt es sich um ein Verfahren der Textextraktion, bei der durch linguistische oder statistische Verfahren Textstellen bzw. einzelne Wörter ermittelt werden, mit deren Hilfe ein Text repräsentiert werden soll.

Man kann zwischen zwei Verfahren, die bei Suchmaschinen Anwendung finden, unterscheiden: zum einem die Darstellung von signifikanten Sätzen und zum anderen einzelne charakteristischen Wörter zur Präsentation eines Dokuments. Beide Varianten werden von den Suchmaschinen in meist Verbindung mit der Hervorhebung der Suchbegriffe verwendet, das heißt, es werden jene Textstellen bzw. Wörter angezeigt, die einen oder mehrere der Suchbegriffe enthalten (Hearst 2006, S. 290). Der Vorteil dieser Verfahren (insbesondere jener Variante bei der Ausschnitte des Textes genutzt werden) ist der, dass der Nutzer einen direkten Einblick in das Dokument erhält und dabei auch ersichtlich werden kann, in welchem Kontext der Suchbegriff in Dokument verwendet wird (Lewandowski 2005, S. 225)

## **Metadaten aus dem Quelltext**

Unter Metadaten versteht man Informationen, die Informationen zu einem Dokument enthalten. Die meisten Dokumente im Internet enthalten Metadaten, die sowohl formale, als auch inhaltliche Angaben zu einem Dokument liefern. In dem bei Webseiten am meisten verwendeten Dateiformat HTML (Hypertext Markup Language) sind diese Metadaten in dem Kopfbereich des Quelltextes enthalten (Erlhofer 2007b, S. 42). Für die inhaltliche Dokumentpräsentation ist besonders der Meta-Tag Description geeignet, da dieser eine kurze Beschreibung des Inhalts bzw. des Angebots einer Webseite enthält (Erlhofer 2007b, S. 45 f.). Eine standardisierte Form, die aber nur vereinzelt Anwendung findet, sind die Meta Daten nach dem Dublin Core Standard. Das entsprechende Feld heißt DC.Description. In der Spezifikation heißt es dazu: „A textual description of the content of the resource, including abstracts in the case of document-like objects or content descriptions in the case of visual resources.“(Dublin Core Metadata Initiative 1998). Im Gegensatz zum KWIC-Ansatz sind diese beschreibenden Metada-

ten nicht zwingend in einem Dokument vorhanden, sondern bedürfen der Initiative des Erstellers bzw. Seitenbetreibers, solche Informationen einem Dokument hinzuzufügen (Haynes 2004, S. 77).

### **Verzeichnisse**

Hierunter sollen solche Verzeichnisse verstanden werden, in denen Webseiten gelistet sind. Diese Einträge enthalten in der Regel allgemein gehaltene Beschreibungen des Inhalts einer Webseite. Als bekanntester Vertreter soll an dieser Stelle das Open Directory Project (ODP) sowie das Yahoo-Verzeichnis<sup>5</sup> genannt werden. Beim ODP handelt es sich um ein Verzeichnis, bei dem freiwillige Editoren die Webseiten erfassen (Open Directory Project 2008a). Das Verzeichnis selbst wird von der Netscape Communication Corporation betrieben. Die Inhalte des ODPs stehen jedoch unter „free license agreement and attribution“, was bedeutet, dass die Inhalte kostenlos und frei Erhältlich, sowie in einem einheitlichen Format verfügbar sind (Open Directory Project 2008b). Dies ist gerade in Bezug auf die Verwendung in Suchmaschinen interessant, da durch die freie Verfügbarkeit den Suchmaschinen ermöglicht wird, diese Beschreibungen zu verwenden und durch das einheitliche Format ein maschineller Zugriff vereinfacht wird. Im Yahoo-Verzeichnis wird die Auswahl und Beschreibung der Webseiten von einem eigenen Redaktionsteam vorgenommen (Yahoo Deutschland 2008). Der Nachteil dieser Verzeichnisse ist darin zu sehen, dass in der Regel nur die Hauptseite und nicht jede Seite einer Webseite beschrieben und erfasst werden.

## **2 Studien zu Trefferbeschreibungen**

Wie schon in der Einleitung thematisiert, wurden in den Studien zu Suchmaschinen die Trefferbeschreibungen selbst kaum berücksichtigt. Im Folgenden soll ein kurzer Überblick über die Studien, die auf die Trefferbe-

---

<sup>5</sup> <http://de.dir.yahoo.com/>

schreibungen eingehen, gegeben werden, sowie zwei Studien näher betrachtet werden.

In bislang 3 Studien werden Trefferbeschreibungen in Untersuchungen zur Retrievaleffektivität der Web-Suchmaschinen berücksichtigt. In der Studie, die 2000 von Wolf durchgeführt wurde, waren es eher pragmatische Gründe, die dazu führten, die Beschreibungen zu berücksichtigen, da das Ausdrucken der einzelnen Treffer und damit die Relevanzbeurteilung der einzelnen Webseiten mit erheblichen Aufwand verbunden gewesen wäre (Wolf 2000). Griesbaum vergleicht in seiner Studie von 2004 die Relevanzbewertungen der Trefferbeschreibungen sowie der Webseiten der Treffer. Dabei stellt er fest, dass bei den untersuchten Suchmaschinen die Beurteilungen in einem Drittel bis zur Hälfte der Fälle unterschiedlich bewertet wurden. In Bezug auf die Veränderung der Beurteilung liegen bei Google und Yahoo gleich viele Fälle vor, bei denen einerseits die Beschreibungen als relevant beurteilt wurden, und die Treffer sich als nicht relevant erwiesen haben, und andererseits der umgekehrten Konstellation, in denen die Trefferbeschreibungen nicht als relevant beurteilt wurden, aber die Webseiten sich als relevant erwiesen. Bei der Suchmaschine Altavista lag der Anteil der Fälle, in denen die Webseiten als relevant beurteilt und die Beschreibungen als nicht relevant befunden wurden, deutlich höher (Griesbaum 2004). Allerdings wurden in dieser Studie die Trefferseiten nicht anonymisiert und die Trefferlisten in der original Reihenfolge belassen, so dass mögliche Effekte aufgrund der Zuordnung zu einer bestimmten Suchmaschine bzw. der Rankingposition nicht auszuschließen sind.

Eine weitere Studie mit einem ähnlichen Untersuchungsgegenstand wurde 2007 von Lewandowski durchgeführt, jedoch geht diese differenzierter auf die Unterschiede der Relevanzbeurteilung bei der Trefferbeschreibungen und der Trefferwebseiten ein (Lewandowski 2008). Auf die Methodik und die Ergebnisse dieser Studie soll weiter unten eingegangen werden.

Einen anderen Untersuchungsfokus setzt die Studie von Crystal und Greenberg aus dem Jahr 2006. Hier wurde untersucht, welche Kriterien bei der

Relevanzbeurteilung der Trefferbeschreibungen und der Trefferdokumente herangezogen werden und in welchen Elementen diese ausschlaggebenden Informationen vorzufinden sind. Zwar beschränkt sich diese Studie auf ein Informationsbedürfnis im Bereich Gesundheitsinformationen, aber sie gibt einen Aufschluss, welche Elemente einer Trefferbeschreibung für die Relevanzbeurteilung auf den Trefferseiten als ausschlaggebend erachtet werden (Crystal u.a. 2006). Auf diese Studie soll ebenfalls genauer eingegangen werden.

In Bezug auf Usability-Aspekte untersucht Kaczmirek 2006 die Trefferseiten und geht dabei auch auf die Trefferbeschreibungen ein. Dabei werden verschiedene Varianten der Trefferbeschreibungen auf verschiedene Kriterien hin untersucht, darunter auch welche Variante zur höchsten Genauigkeit oder zur höchsten Vollständigkeit führt. Weiterhin gehört zu den Untersuchungsfragen, welche Beschreibungsvariante im direkten Vergleich von den Nutzern bevorzugt wird (Kaczmirek 2006, S. 60). Zur Auswahl standen folgende Varianten: a) Beschreibungen nach dem KWIC-Ansatz, b) Ausschnitt aus dem Anfang des Dokuments<sup>6</sup>, c) die Extraktion der 7 semantisch sinnvollen Substantive d) einer Kombination der Varianten a-c (Kaczmirek 2006, S. 44 f.). Kaczmirek kommt zum Ergebnis, dass in Bezug auf Genauigkeit und Vollständigkeit zwischen den verschiedenen Beschreibungsvarianten keine signifikanten Unterschiede festzustellen sind. In Bezug auf die Untersuchungsfrage, welche Variante zur höchsten Erfolgsrate führt, wurde festgestellt, dass die Variante, mit einzelnen signifikanten Wörtern (c) im Vergleich anderen drei Varianten höher ausfällt. Weiterhin wurde ein besseres Abschneiden der Variante Wortanfang, im Vergleich zu jener Variante in der alle Varianten gemischt werden festgestellt. Im direkten Vergleich der Beschreibungsvarianten wurden Beschreibungen, die den Anfang des Textes enthalten sind, den anderen Beschreibungsmöglichkeiten vorge-

---

<sup>6</sup> Hierbei handelt es sich um ein von Kaczmirek vorgeschlagenes Verfahren, bei dem der erste Bereich mit einem zusammenhängenden Text identifiziert wird und der Ausschnitt im Gegensatz zu den meist üblichen 200 Zeichen, 400 Zeichen beträgt.

zogen. Abschließend wird diese auch als Favorit für die Trefferbeschreibung gesehen, wobei bei spezifischen Suchanfragen diese mit der KWIC-Methode gleichrangig gesehen wird (Kaczmirek 2006, S. 80 ff.).

## 2.1 Studie Lewandowski 2008<sup>7</sup>

In dieser bereits weiter oben erwähnten Studie untersucht Lewandowski neben den Relevanzbeurteilungen der Webseiten auch die der Trefferbeschreibungen. Untersucht wurden die Webseiten als auch die Beschreibungen in insgesamt 40 Suchanfragen die ersten 20 Treffer der Suchmaschinen Google, Yahoo, MSN, Ask sowie Seekport. Im Gegensatz zu der Studie von Griesbaum wurden auch die Trefferbeschreibungen anonymisiert und nach einem Zufallsprinzip angeordnet, so dass keine Rückschlüsse auf die Suchmaschine oder die Rankingpositionen möglich waren. Der Bewertungsprozess wurde in zwei Phasen durchgeführt: Zuerst wurden die Trefferbeschreibungen und anschließend die Trefferwebseiten bewertet. Damit wurden auch jene möglichen Effekte verhindert, die bei Griesbaum durch die Bewertung der Trefferbeschreibungen und der Webseiten durch unterschiedliche Personen hervorgerufen werden können. Diese Studie kommt im bezogen auf die Ergebnis Webseiten je nach Suchmaschine zu einem Precision zwischen 0,37 und 0,52 und bei den Trefferbeschreibungen auf Werte zwischen 0,44 und 0,60. Die Abweichung zwischen den beiden Precision-Werten ist bei Google und MSN etwas höher, als bei den drei anderen Suchmaschinen. Weiterhin betrachtet Lewandowski, wohin sowohl die als relevant, als auch die nicht relevant beurteilten Treffer führen. Die relevanten Beschreibungen, die zu ebenso relevanten Treffern führen, werden als „Description-result Precision“ gemessen, der zwischen 0,27 und 0,41 liegt. Unter „description-result confidence“ wird betrachtet, in wie vielen Fällen sowohl die in relevanten, als auch nicht relevanten Beschreibungen mit der Beurteilung der Webseiten übereinstimmen (0,70-0,79). Der „description fallout“ stellt dar, wie oft relevante Ergebnisse nicht als relevant erkannt

---

<sup>7</sup>Lewandowski 2008

werden würden, da die zugehörigen Trefferbeschreibungen als nicht relevant eingestuft wurden (0,07-0,11). „Description deception“ betrachtet den gegenteiligen Fall, also in wie vielen Fällen eine als relevant betrachtete Beschreibung zu einem als relevant bewerteten Ergebnis führt (0,14-0,21). Während die Precision-Werte, bei denen sowohl der Treffer, als auch die Beschreibung als relevant eingestuft in einem unbefriedigenden Niveau liegen, kann festgestellt werden, dass in 70% bis 79% der Fälle eine Übereinstimmung zwischen der Trefferbeschreibung und dem Treffer vorliegt. Damit liegen diese insgesamt in einem etwas höheren Bereich, als es von Griesbaum 2004 festgestellt wurde (63-65%). Dennoch werden im Schnitt ein Viertel der Fälle die Trefferbeschreibung anders als das Dokument selbst bewertet, so dass nicht zwingend Rückschlüsse auf die Qualität eines Dokuments anhand einer Trefferbeschreibung gegeben sind. Betrachtet man jene Treffer, deren Beschreibung und Dokument uneinheitlich bewertet wurden, so kann festgestellt werden, dass der Anteil der als relevant betrachteten Beschreibungen (die zu einem als nicht relevant betrachteten Treffer führen) höher liegt, als denjenigen, bei denen die Beschreibungen als nicht relevant beurteilt wurden (deren die Treffer jedoch relevant waren).

## **2.2 Crystal und Greenberg 2006<sup>8</sup>**

Bei dieser Studie lag der Hauptfokus nicht im Bezug auf die Retrievaleffektivität, sondern darauf welche Elemente einer Trefferbeschreibung bzw. eines Treffers für die Relevanzbeurteilung als ausschlaggebend gesehen werden.

Im Rahmen dieser Studie führten die Teilnehmer eine Recherche in einer Websuchmaschine zu einer selbstdefinierten Fragestellung im Bereich Gesundheitsinformationen durch. Während dieser Recherche wurden die Teilnehmer dazu aufgefordert, die für die Beurteilung relevanten Informationen zu markieren. In einem weiteren Schritt wurden die Bildschirmaufzeichnun-

---

<sup>8</sup> Crystal et al. 2006

gen mit der Testperson betrachtet und die Testperson sollte ihre Überlegungen bei der Recherche entsprechend der „Thinking-Aloud-Method“ kommentieren. Die Auswertung erfolgte in 2 Phasen. In der ersten wurden alle zur Urteilsfindung als relevant eingestuften Kriterien erfasst. Weiterhin wurden diese Kriterien lokalisiert und einem Bereich, wie z.B. Titel, Beschreibung, URL etc. zugeordnet. In der zweiten Phase wurden diese nach Häufigkeit gebündelt, um Aufschluss zu erhalten, welche Kriterien sowie welche Bereiche eines Dokuments bzw. einer Beschreibung besonders häufig zur Relevanzbeurteilung herangezogen werden. Bemerkenswert ist die Tatsache, dass sich über 90% der für den Nutzer relevanten Merkmale in der Beschreibung selbst befanden. In Bezug darauf, welche Kriterien als besonders wichtig gesehen werden, kommen Greenberg und Crystal zum Ergebnis, dass sowohl bei den Trefferbeschreibungen als auch auf den Webseiten neben den an erster Stelle stehenden thematischen Kriterien eine Vielzahl von weiteren Kriterien herangezogen werden, um ein Urteil zu fällen. Dabei wurden an zweiter Stelle, nach den Angaben zum Verfasser, jene Kriterien identifiziert, die dazu dienen, irrelevante Treffer zu identifizieren bzw. herauszufiltern.

Bezogen auf die Konsistenz der Relevanzbeurteilungen konnte festgestellt werden, dass von jenen Dokumenten, deren Beschreibungen als relevant erachtet wurden 57,4% ebenfalls als relevant beurteilt wurden. Ein anderes Bild geben jene Treffer, deren Beschreibungen als nicht relevant beurteilt wurden. Hier zeigte sich, dass entgegen der ersten Bewertung 63,2% der Beschreibungen als relevant befunden wurden.

### **2.3 Zusammenfassende Würdigung der Studien**

Während in den Studien von Griesbaum und Lewandowski die Trefferbeschreibungen in Bezug auf den Relevanzbeurteilungsprozess betrachtet wurden und die Differenzen zwischen der Beschreibung und dem Dokument im Vordergrund standen, beschäftigten sich Crystal und Greenberg primär mit den Kriterien und den Elementen einer Beschreibung, die zu einer Rele-

vanzaussage führen. Kaczmirek dagegen untersucht die verschiedenen Konzepte von Trefferbeschreibungen und wie diese sich auf die Beurteilung der Treffer auswirken. Zusammenfassend kann gesagt werden, dass eine hohe Anzahl der Beschreibungen übereinstimmend mit den Trefferwebseiten eingeschätzt werden. Jedoch zeigt insbesondere die Studie von Lewandowski, dass der Anteil jener Treffer, in denen als relevant eingestufte Beschreibungen auf ebenfalls relevante Treffer verweisen, relativ gering ist. Crystal und Greenberg zeigen, dass die Nutzer ihr Urteil anhand einer Vielzahl von Kriterien fällen. Hierbei stellt sich die Frage, mit welchen Informationen die Beschreibungen aus den Dokumenten zusätzlich zu den bestehenden Elementen einer Trefferbeschreibung zu ergänzen sind, um den Nutzer bei seiner Entscheidungsfindung zu unterstützen. In der von Kaczmirek durchgeführten Studie wird deutlich, dass die verschiedenen Beschreibungsvarianten unterschiedlich gut geeignet sind, den Nutzer bei dem Evaluierungsprozess zu unterstützen. Allerdings basiert diese Untersuchung nur auf einer relativ kleinen Auswahl verschiedener Trefferbeschreibungen, so dass offen bleibt, wie die verschiedenen Beschreibungstypen in einem größeren Spektrum und in Bezug auf verschiedene Suchmaschinen den Nutzer bei dem Entscheidungsprozess unterstützen.

### **3 Untersuchung der Trefferbeschreibungen und deren Bewertung**

Die zuvor vorgestellten Studien zeigen, dass den Trefferbeschreibungen selbst und ihren strukturellen und qualitativen Unterschieden noch kaum Aufmerksamkeit geschenkt wurde. Dies ist sofern erstaunlich, da gerade die Beschreibungen ausschlaggebende Informationen geben können, welche Treffer dazu geeignet sind, das Informationsbedürfnis zu befriedigen. Wie in Abbildung 3 zu erkennen ist, fallen die Beschreibungen für dieselbe Webseite in den einzelnen Suchmaschinen teilweise sehr unterschiedlich aus. Selbst dann, wenn auf die gleiche Beschreibungsquelle zurückgegriffen wird, unterscheiden sich diese bezüglich der Länge der Ausschnitte, sowie

in der Struktur. So stellt sich die Frage, wie häufig die verschiedenen Beschreibungsvarianten bzw. -quellen von den Suchmaschinen genutzt und wie diese von den Nutzern angenommen werden. Im Folgenden soll eine vom Verfasser durchgeführte Untersuchung vorgestellt werden, in der die Struktur der Trefferbeschreibungen und deren Bewertungen analysiert werden.

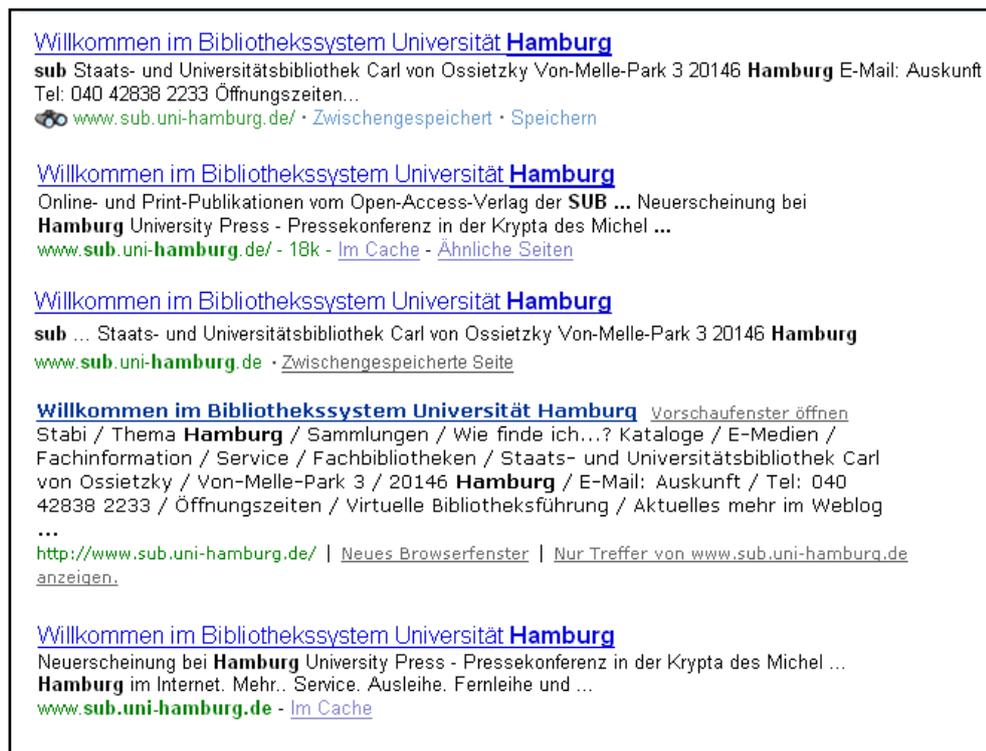


Abbildung 3 : von oben nach unten: Ask, Google, MSN/Live, Seekport, Yahoo

### 3.1 Fragestellungen

Ausgehend von der Erkenntnis, dass die Suchmaschinen unterschiedliche Konzepte und Methoden für die Erzeugung der Trefferbeschreibung anwenden und diese auch unterschiedlich ausfallen, stellt sich die Frage, wie verbreitet die einzelnen Methoden sind und wie gut die verschiedenen Methoden dazu geeignet sind Inhalte zu repräsentieren. Weiterhin stellt sich die Frage, inwiefern sich bestimmte Merkmale, wie zum Beispiel die Länge einer Beschreibung oder die sprachliche Qualität, auf die Beschreibungen in ihrer Funktion als aussagekräftiger Repräsentant des Inhalts eines Dokuments auswirkt.

Daraus ergeben sich folgende Fragen, die den Kern der Untersuchung bilden:

- F1. Wie häufig werden die verschiedenen Beschreibungsquellen von den Suchmaschinen genutzt?
- F2. Welche Merkmale haben die Beschreibungen bezüglich ihrer Länge und sprachlicher Qualität?
- F3. Wie werden die Beschreibungsquellen von den Nutzern angenommen?
- F4. Welche Merkmale haben einen Einfluss auf die Bewertung der Beschreibungen?

### **3.2 Methodik**

Die Studie untergliedert sich in zwei Abschnitte: die Strukturanalyse der Trefferbeschreibungen sowie eine Befragung mit Fragebogen, in der die Beschreibungen bewertet werden sollen. Die Grundlage dieser Untersuchung bilden die ersten 10 Trefferbeschreibungen zu 100 Suchanfragen, die an fünf Suchmaschinen gestellt werden. Es handelt es sich dabei um die vier größten Suchmaschinen Google, Yahoo, MSN/Live<sup>9</sup> und Ask (Webhits 2008); (Comscore 2008). Als weitere Suchmaschine wird auch Seekport berücksichtigt, da diese die einzige deutsche Suchmaschine mit eigener Technologie und eigenem Index ist (at-web 2004; Alby u.a. 2007, S. 40 ff.).

#### **Erhebung der Daten**

Die 100 Suchanfragen werden als Zufallsstichprobe den Top 1 Millionen Suchanfragen der Suchmaschine T-Online von März 2007 entnommen. Betrachtet man die Länge der Suchanfragen, so sind von den 102 Anfragen<sup>10</sup> 81, die aus einem Wort bestehen, sowie 21 Formulierungen, die aus 2

---

<sup>9</sup>Diese Suchmaschine der Firma Microsoft wird seit einiger Zeit unter dem Namen Live angeboten.

<sup>10</sup> Die 100 Suchanfragen wurden um 2 weitere ergänzt, um einerseits Reserve zu erhalten und andererseits fehlende Treffer auszugleichen.

Wörter bestehen. Damit liegt der Schnitt bei 1,21 Wörtern pro Suchanfrage und ist somit etwas kürzer als der allgemeine Durchschnitt, der bei 1,7 Wörtern liegt (Schmidt-Mänz 2007, S. 129 f.). Auf eine Beschränkung der Suchanfragen auf einen bestimmten Anfragetyp wird verzichtet, da im Fokus dieser Studie die Beurteilung der Trefferbeschreibungen steht und nicht beispielsweise Entscheidungsmuster, die darüber Auskunft geben, welche Trefferbeschreibung als erstes selektiert werden. In einem weiteren Schritt wird die Stichprobe einerseits von Doubletten und andererseits von jenen Suchanfragen bereinigt, die auf eine sexuelle bzw. pornographische Motivation der Suche hindeuten, da bezweifelt werden kann, dass die Trefferbeschreibungen solcher Anfragen bei der Bewertung mit der entsprechenden Ernsthaftigkeit beurteilt werden.

Diese Suchanfragen werden zeitgleich an die fünf Suchmaschinen gestellt und jeweils die erste Trefferseite gespeichert. Damit ergibt sich eine Beschränkung auf die ersten 10 Treffer. Dies spiegelt auch das typische Verhalten der Nutzer wider, da diverse Studien gezeigt haben, dass in den meisten Fällen ohnehin nur die erste Trefferseite berücksichtigt wird (Machill u.a. 2003, S. 255; Erlhofer 2007a, S. 156 ff.).

Um die Zahl der vergleichbaren Ergebnisse hinsichtlich der Sprache zu erhöhen, wird bei den Suchmaschinen die Einstellung für eine Begrenzung auf deutschsprachige Seiten verwendet, sofern dies nicht ohnehin der Standardeinstellung entspricht. Die Treffer werden in einer Datenbank erfasst und auf ihre Merkmale wie Länge oder die sprachliche Beschaffenheit hin untersucht. Diese erfassten Treffer bilden gleichzeitig die Grundlage für die Fragebögen, in denen die Beschreibungen bewertet werden sollen.

### **Zu erfassende Merkmale**

Neben dem Inhalt eines Treffers bzw. einer Trefferbeschreibung, der aus dem Titel, der Beschreibung und der URL besteht, werden noch weitere Merkmale erfasst. Dazu gehören: Die Suchmaschine, die Suchanfrage, zu der der Treffer gehört, sowie die Position des Treffers innerhalb der erfassten Trefferseite. Bezüglich der Beschreibungsquellen ergeben sich bezug-

nehmend auf den Abschnitt 1.4 folgende Ausprägungen: Text aus dem Seiteninhalt selbst (KWIC), Metadaten, Verzeichnisse (hier nach ODP und Yahoo-Verzeichnis getrennt), sowie eine weitere Ausprägung „Sonstiges“, in der jene Beschreibungen erfasst werden können, die nicht eindeutig einem der oben genannten Beschreibungsquellen zugeordnet werden können. Für die sprachliche Beschaffenheit der Beschreibungen wird zwischen folgenden Ausprägungen differenziert: vollständiger Satz, Ausschnitt aus einem Satz, einzelne Wörter, sowie fremdsprachige Beschreibungen. Die Bewertung durch die Teilnehmer der Befragung wird auf einer Skala von 1 bis 6 vorgenommen. Die Skalenwerte entsprechen den deutschen Schulnoten, das bedeutet, dass die 1 dem „sehr gut“ und die 6 „ungenügend“ entspricht. Weitere Merkmale einer Trefferbeschreibung, wie zum Beispiel die Länge einer Beschreibung, werden von der Datenbank automatisch ermittelt.

### **Fragebogen**

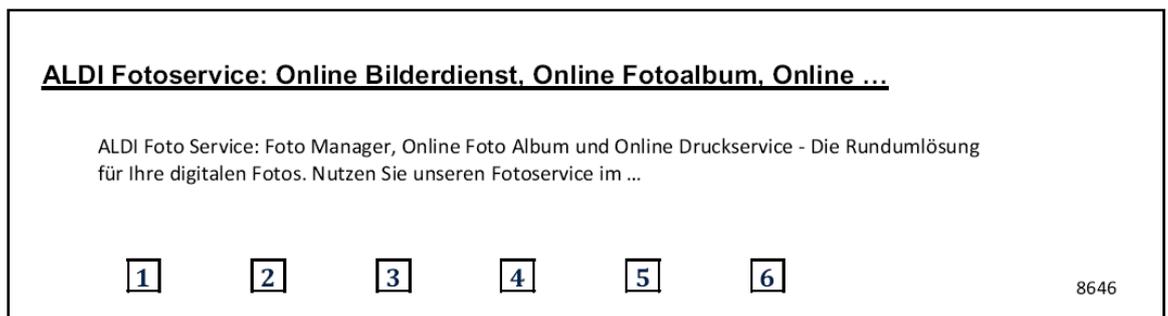
Ziel der Bewertung der Trefferbeschreibungen ist es, Erkenntnisse zu gewinnen, wie die Beschreibungen in Bezug auf ihrer Aussagekraft des Inhalts einer Trefferwebseite eingeschätzt werden. Das heißt, die Teilnehmer sollten bewerten, wie aussagekräftig die Trefferbeschreibung den zu erwartenden Inhalt einer Webseite beschreibt. Den Schwerpunkt der Bewertung auf die Aussagekraft zu legen, erschien aus mehreren Gründen sinnvoll: zum einem sollte dadurch der repräsentative Charakter, den eine Trefferbeschreibung hat, hervorgehoben werden und zum anderen somit vermieden werden, dass die Relevanzbeurteilung eines Treffers vorgenommen wird.

Die Fragebögen werden unter der Berücksichtigung der nachfolgenden Aspekte mithilfe der Datenbank generiert. Um Rückschlüsse auf die Suchmaschinen zu vermeiden, werden die Trefferlisten der einzelnen Suchmaschinen neutralisiert und in einem weiteren Schritt nach einem Zufallsprinzip durchmischt, so dass keine Rückschlüsse auf die Suchmaschine, sowie die Rankingposition eines Treffers mehr möglich sein sollten. Die einzelnen Treffer enthalten den Titel eines Treffers und die dazugehörige Beschreibung. Auf die Angabe der URL, die ebenfalls standardmäßig zur Trefferbe-

beschreibung gehört(vgl. 1.3), wurde verzichtet, da das Risiko besteht, dass die Teilnehmer anhand der URL zu einer Relevanzbeurteilung neigen könnten. Aus demselben Grund wird auch auf eine Nennung des zu einer Treffergruppe gehörenden Suchbegriffs verzichtet.

Um Ermüdungseffekten entgegenzutreten wurde die Option, dass ein Teilnehmer alle Treffer zu einer Suchanfrage bewertet, verworfen und stattdessen eine Variante vorgezogen, bei der die Befragten statt der 50 Trefferbeschreibungen zu einer Suchanfrage jeweils 10 Trefferbeschreibungen zu 5 verschiedenen Suchanfragen erhalten. Ein weiterer Aspekt, den es zu berücksichtigen galt, war die Anzahl der Trefferbeschreibungen pro Seite. Hierbei wurde ebenfalls die zu erst angedachte Variante 10 Trefferbeschreibungen auf einer Seite zu bündeln verworfen, da Trefferbeschreibungen zu dicht aufeinander folgten, und die Gefahr bestand, dass die Trefferbeschreibungen nicht mehr individuell bewertet werden.

Unterhalb des jeweiligen Treffers sind die Skalenwerte für die Bewertung in



**Abbildung 4 Treffer im Fragebogen**

Kästchen, den entsprechenden Wert angeordnet. Für eine solche Anordnung, bei der die Teilnehmer den Wert ankreuzen können, sprach einerseits die Tatsache, dass die volle Skalenbreite sichtbar ist und zum anderen, dass die abgegebene Bewertung eindeutig einem Wert zuzuordnen ist. Um zurückverfolgen zu können, um welchen Treffer es sich handelt, wurde eine Identifizierungsnummer eingebunden. Abbildung 4 zeigt einen Treffer im Fragebogen.

### 3.3 Durchführung Analyse der Trefferbeschreibungen

Die Erfassung der Trefferbeschreibungen wurde im Zeitraum zwischen dem 14. März und dem 28. März 2008 durchgeführt. Diese Zeitspanne resultiert aus der Tatsache, dass die Trefferbeschreibungen nicht nur erfasst, sondern auch in dieser Phase auf den Ursprung der Trefferbeschreibung hin untersucht wurden. Die Trefferbeschreibungen wurden in einer vom Verfasser entwickelten Access-Datenbank erfasst. Berücksichtigt wurden nur jene Treffer, die der Websuche zugeordnet werden konnten, nicht berücksichtigt wurden zum Beispiel bei den normalen Treffern eingebundene Ergebnisse der News-Suche oder andere eingebundene Kollektionen bzw. Suchräume.

Bei der Erfassung der Beschreibungsquellen wurden diese mit den Webseiten abgeglichen, um die Quelle der Beschreibung zu lokalisieren. In jenen Fällen, in denen der Inhalt aufgrund des dynamischen Charakters von Webseiten im Internet nicht mehr vorzufinden war, wurde zuerst mithilfe der Cache-Version der Suchmaschinen, bzw. der in der „Way-Back-Machine“ des Internet-Archive<sup>11</sup> hinterlegten Version versucht zu ermitteln, ob und wo der Inhalt der Beschreibung vorzufinden war. Zwischen den einzelnen Beschreibungsquellen, aus denen sich gleichzeitig die verschiedenen Beschreibungstypen ergeben, wurde folgendermaßen differenziert:

Als Beschreibungen, die aus dem Inhalt der Webseite stammen, wurden jene Beschreibungen eingestuft, deren Text optisch sichtbar auf der Webseite vorzufinden war. Dies ist auch dann der Fall, wenn der Beschreibungstext erst durch eine Interaktion (z.B. das Öffnen eines Menüs) sichtbar wurde. Nicht berücksichtigt wurden jene Inhalte, die entweder nur als Alternativtext für Bilder im Quelltext vorhanden waren, oder jene Inhalte, die man nur dann zu sehen bekommt, wenn Erweiterungen, die heute zum Standard gehören, nicht im Browser integriert sind. So z. B. die Hinweistexte, dass Flash installiert werden soll, oder Java-Script zu aktivieren ist, um den Inhalt einer Seite zu sehen. Solche Fälle wurden unter „Sonstiges“ erfasst.

---

<sup>11</sup> <http://www.archive.org/web/web.php>

Unter Metadaten als Beschreibungsquelle wurden nur jene Beschreibungen aufgenommen, dessen Inhalt im Quelltext auch mit entsprechenden Tags als „beschreibende“ Metadaten ausgewiesen ist (Keywords od. Description). Hierbei wurden jene Fälle nicht berücksichtigt, bei denen der Beschreibungsinhalt zwar im Quelltext eines Dokuments vorzufinden war, jedoch diese nicht durch entsprechende Tags als Metadaten ausgewiesen wurden. Solche Fälle wurden ebenfalls unter Sonstiges erfasst. Beschreibungen, deren Inhalt sowohl im Text einer Seite, als auch in den Metadaten vorzufinden sind, werden immer dem ersteren zugerechnet. Dies ist z.B. dann der Fall, wenn im Quelltext einer Seite, die von einem Content Management System (CMS) generiert wurde, im Description-Tag auf den gesamten Seiteninhalt verwiesen wird, oder wie im Fall von Wikipedia<sup>12</sup> auf eine RDF<sup>13</sup>-Datei verwiesen wird, jedoch deren Beschreibungsinhalte immer mit dem des Textinhalts deckungsgleich ist.

Bei den Verzeichnissen wurde nach getrennt Open Directory Project und Yahoo Verzeichnis.

In einem weiteren Schritt wurde die sprachliche Beschaffenheit der Beschreibungen analysiert und folgenden Kategorien zugeordnet:

- Vollständiger Satz
- Ausschnitt aus einem Satz
- Einzelne Wörter inklusive der Ansammlung von Wörtern bei denen es nicht erkennbar ist, dass die für einen Satz notwendige logische semantische Verbindung der Wörter vorhanden ist
- Fremdsprachige Trefferbeschreibungen, eingeschlossen jener Trefferbeschreibungen, die nicht lesbare Zeichensätze enthalten

---

<sup>12</sup> <http://www.wikipedia.de>

<sup>13</sup> Resource Description File

### **3.4 Durchführung der Bewertung mit Fragebögen**

Den Kern des Fragebogens stellte wie schon in Abschnitt 3.2 dargestellt der Output der Datenbank dar, der die anonymisierten und zufällig angeordneten Beschreibungen enthält. Die Fragebögen wurden zu jeweils aus fünf Suchanfragen bestehenden Sets zusammengestellt. Die genaue Zusammensetzung der einzelnen Sets kann dem Anhang A3 entnommen werden. Dieser Kern des Fragebogens wurde noch um eine Einleitung zu Beginn des jeweiligen Fragebogensets, sowie am Ende mit Fragen zum Suchverhalten ergänzt. Die Einführung enthält zum Nachlesen die wichtigsten Informationen, die auch beim Briefing jeder Befragung mündlich erwähnt wurden. Dazu gehören die genaue Fragestellung der Bewertung, worauf während der Bewertung besonders zu achten ist, ein Beispiel einer Beschreibung, sowie die Darlegung der Notenwerte.

Die Fragen zum Suchverhalten, die den Fragebogen abschließen, dienen dazu, um zu überprüfen, ob die Einflussfaktoren der Bewertungen, wie zum Beispiel sprachliche Beschaffenheit eines Treffers, oder die Länge eines Treffers, mit den erklärten Vorzügen des Teilnehmers übereinstimmen. Gefragt wurde nach der bevorzugten Suchmaschine, was den Teilnehmern bei der Trefferbeschreibung besonders wichtig ist, sowie die Länge der idealen Trefferbeschreibung. Ein Beispielfragebogen ist dem Anhang A2 zu entnehmen.

Die Bewertungen wurden in insgesamt fünf Seminaren durchgeführt: Drei Seminaren des 2. Semesters im Studiengang Medien und Information (insges. 51 TN), einem Seminar der Studiengänge Medien- und Information sowie Bibliotheks- und Informationsmanagement (ab 3. Semester, 27 TN) sowie 17 Studierenden aus dem Diplom-Studiengang „Technische BWL“ der Hochschule für Angewandte Wissenschaften Hamburg (HAW) sowie zwei weiteren Studierenden aus weiteren Fachrichtungen durchgeführt.

### 3.5 Ergebnisse

In den folgenden Abschnitten sollen die Ergebnisse der Untersuchungen betrachtet und in Bezug auf die Fragestellungen diskutiert werden.

#### 3.5.1 Analyse der Trefferbeschreibungen

Die Auswertung der Trefferbeschreibungen zu den 102 Suchanfragen ergab eine Gesamtanzahl von 5027 Trefferbeschreibungen. Die Differenz von 73 fehlenden Beschreibungen zum Soll-Wert von 5100 Beschreibungen ergibt sich dadurch, dass nicht alle Suchmaschinen in der Lage waren mindestens 10 Treffer zu einer Suchanfrage zu liefern. Tabelle 1 zeigt die Anzahl der Treffer der jeweiligen Suchmaschine sowie die Anzahl der fehlenden Treffer, sowie die Anzahl der Anfragen, die keinen einzigen Treffer lieferten. Während Google, Ask und Yahoo stets mindestens 1 Treffer zu einer Suchanfrage liefern konnten, lieferten Seekport und MSN in einem bzw. mehreren Fällen nicht einmal mindestens einen Treffer.

|                       | Ask   | Google | MSN/Live | Seekport | Yahoo  |
|-----------------------|-------|--------|----------|----------|--------|
| Anzahl                | 1019  | 1020   | 982      | 1002     | 1004   |
| Fehlend <sup>14</sup> | 1/1/0 | 0/0/0  | 38/4/3   | 18/3/1   | 16/2/0 |

**Tabelle 1 Anzahl der Treffer**

Die Untersuchung der Beschreibungsquellen bzw. Beschreibungstypen ergab, dass der KWIC Ansatz am häufigsten verwendet wird. Der Anteil liegt zwischen etwas unter 36% bei MSN und etwas über 89% bei Seekport. Hierbei ist anzumerken, dass Seekport etwas aus dem Rahmen fällt, da diese nur auf eine Beschreibungsquelle zu setzen scheint. In Bezug auf alle Trefferbeschreibungen liegt dieser bei 51%.

---

<sup>14</sup> Die erste Ziffer gibt an, wie viele Treffer fehlen und die zweite, wie viele Suchanfragen nicht mindestens 10 Treffer enthielten. Die letzte Ziffer gibt die Anzahl der Anfragen, in denen kein Treffer geliefert wurde, an

An zweiter Stelle folgt die Verwendung von Metadaten als Beschreibungsquelle. Der Anteil verteilt sich zwischen 37% bei Ask und knapp 52% bei MSN. Eine Ausnahme stellt Seekport dar, dessen Anteil aufgrund des hohen Anteils an KWIC-Quellen bei 1,9% liegt.

| Typ                     | Ask            | Google         | MSN/Live       | Seekport       | Yahoo          | Alle SM        |
|-------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| KWIC                    | 45,44%         | 45,78%         | 35,85%         | 89,12%         | 37,95%         | <b>51,02%</b>  |
| Metadaten               | 37,78%         | 43,14%         | 51,93%         | 1,90%          | 48,31%         | <b>36,42%</b>  |
| ODP                     | 9,91%          | 6,67%          | 8,15%          | 0,00%          | 1,00%          | <b>5,13%</b>   |
| Sonstiges <sup>15</sup> | 5,99%          | 3,04%          | 2,95%          | 7,49%          | 6,18%          | <b>5,13%</b>   |
| Ohne Besch.             | 0,88%          | 1,37%          | 1,12%          | 1,50%          | 6,57%          | <b>2,29%</b>   |
| <b>Summe</b>            | <b>100,00%</b> | <b>100,00%</b> | <b>100,00%</b> | <b>100,00%</b> | <b>100,00%</b> | <b>100,00%</b> |
| N                       | 1019           | 1020           | 982            | 1002           | 1004           | 5027           |

**Tabelle 2** Anteile der Beschreibungstypen

Bezogen auf das Open Directory Project als Beschreibungsquelle, liegt der Anteil bei Ask, Google und MSN zwischen 6,57% und 9,91%. Dagegen finden sich bei Seekport keine Beschreibungen, deren Ursprung beim ODP liegen. Bei Yahoo belaufen sich diese auf 1%. Es wurde angenommen, dass Yahoo auf sein eigenes Verzeichnis zurückgreift, aber eine solche Beschreibungsquelle konnte nicht nachgewiesen werden. Dies ist jedoch mit den aktuellen Äußerungen seitens der Entwicklung von Yahoo in Einklang zu bringen, da Yahoo auf neuere Ansätze setzt als ein von Menschen manuell geführtes Verzeichnis. In einem Interview äußert Andreas Krawczyk, Leiter der Programmierung bei Yahoo Deutschland, es folgendermaßen:

*„In Zeiten der algorithmischen und semantischen Suche und zugleich zig Millionen von Webseiten gibt es praktikablere Methoden, dem Nutzer eine schnelle relevante Antwort zu seiner Recherche zu geben.“* (at-web 2008b).

<sup>15</sup> Mit enthalten sind jene Beschreibungen, die auf technische Probleme seitens der Suchmaschine hindeuten. Die Anteile bezogen auf die unter Sonstiges erfassten Beschreibungen:

| Ask    | Google | MSN/Live | Seekport | Yahoo  | Summe  |
|--------|--------|----------|----------|--------|--------|
| 21,31% | 16,13% | 24,14%   | 10,67%   | 17,74% | 17,05% |

Der Anteil jener Beschreibungen, die entweder nicht zugeordnet werden konnten, oder Hinweise enthalten, dass die Suchmaschine Probleme hatte den eigentlichen Inhalt einer Webseite zu erfassen, wie z.B. Seiten, bei denen Java Script o.ä. aktiviert sein muss, wurden unter Sonstiges erfasst. Der Anteil hierbei liegt zwischen 2,95% und 7,49%. Davon entfallen zwischen 10,67% und 24% auf solche technisch bedingten Beschreibungen. Tabelle 2 gibt einen detaillierten Überblick über die Anteile der Beschreibungstypen bzw. -quellen.



**Abbildung 5 Mehrere Elemente: hier Yahoo**

Während der der Analyse wurde deutlich, dass die Trefferbeschreibungen nicht nur auf eine Beschreibungsquelle zurückgreifen, sondern häufig mehrere Auszüge einbinden, die entweder auf verschiedene oder auf die gleiche Beschreibungsquelle basieren.

Charakteristisch für die Verwendung mehrere Elemente in einer Trefferbeschreibung ist das „...“ zwischen zwei Abschnitten innerhalb einer Trefferbeschreibung. Abbildung 5 verdeutlicht die Einbindung mehrerer Extrakte eines Treffers in einer Trefferbeschreibung. Im Weiteren soll bei diesen Ausschnitten von Elementen gesprochen werden.

Tabelle 3 gibt Auskunft darüber, wie häufig wie viele Elemente vorkommen. Dabei wird deutlich, dass bei Ask, Yahoo und MSN die meisten Beschreibungen aus einem einzigen Ausschnitt bestehen, und der Anteil jener Beschreibungen, die aus mindestens 2 Elementen besteht, unter einem Viertel der Beschreibungen beträgt. Bei Yahoo entgegen enthalten etwas mehr als drei Viertel der Beschreibungen mindestens 2 Elemente, sowie knapp einem Viertel aus einem einzigen Element. Bei Seekport dagegen verteilt es

sich auf einer breiten Bandbreite, die sich auf bis zu mehr als 11 Elemente beläuft. Allerdings enthalten knapp über zwei Drittel maximal 3 Elemente. Betrachtet man die Beschreibungen, die mehrere Elemente enthalten in Zusammenhang mit den Beschreibungsquellen, so ergibt sich folgende Zusammensetzung: Auf jene Beschreibungen, deren erstes Element dem KWIC zugeordnet sind, entfallen insgesamt knapp 65%. Auf Metadaten entfallen knapp 30%. Weiterhin sind unter einem Prozent dem ODP zugeordnet und zu den unter Sonstiges erfassten Beschreibungen knapp über 5,09%. In einer weiteren Untersuchung wurden die jeweils zweiten Elemente einer mehrteiligen Trefferbeschreibung einer Beschreibungsquelle zugeordnet. Daraus ergeben sich folgende Paarungen, die Abbildung 6 und Abbildung 7 entnommen werden können.

| Elemente   | Ask    | Google | MSN/Live | Seekport | Yahoo  | Alle SM |
|------------|--------|--------|----------|----------|--------|---------|
| 1          | 82,67% | 85,29% | 77,96%   | 21,28%   | 24,63% | 58,86%  |
| 2          | 16,93% | 13,32% | 21,94%   | 25,23%   | 64,93% | 28,01%  |
| 3          | 0,30%  | 1,29%  | 0,10%    | 19,05%   | 10,23% | 6,13%   |
| 4          | 0,10%  | 0,10%  | 0,00%    | 15,20%   | 0,21%  | 3,14%   |
| 5          | 0,00%  | 0,00%  | 0,00%    | 7,80%    | 0,00%  | 1,57%   |
| =>6        | 0,00%  | 0,00%  | 0,00%    | 11,45%   | 0,00%  | 2,30%   |
| Anteil =>2 | 17,33% | 14,71% | 22,04%   | 78,72%   | 75,37% | 41,14%  |

**Tabelle 3 Anzahl der Elemente je Suchmaschine**

Dabei ist festzustellen, dass bei allen Konstellationen, jene Beschreibungen dominieren, die im zweiten Element KWIC zugeordnet sind. Kumuliert man all jene Treffer, in denen entweder im ersten oder zweiten Element der KWIC-Methode zugeordnet werden, so kommt man auf einen Anteil von knapp 98%. Berücksichtigt man alle Trefferbeschreibungen, also auch jene, die nur eine einzige Beschreibung enthalten, so kommt der KWIC Ansatz in Bezug auf alle Trefferbeschreibungen auf einen Anteil von 65,61 % (vgl. Anlage A4).

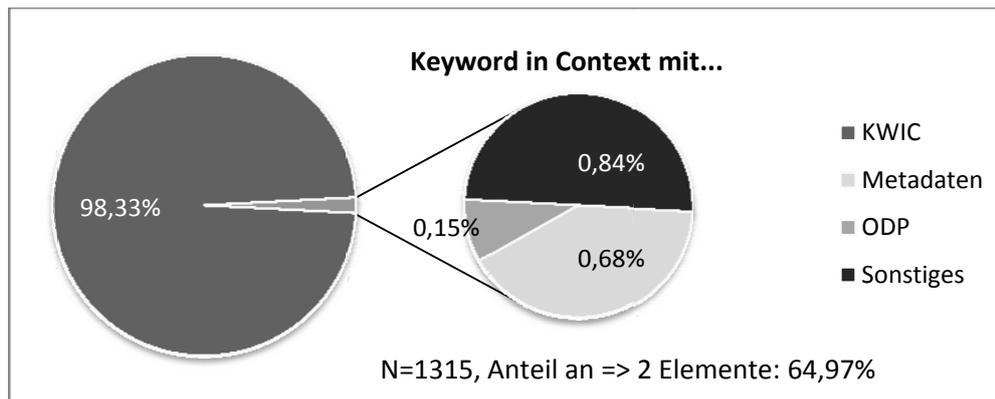


Abbildung 6

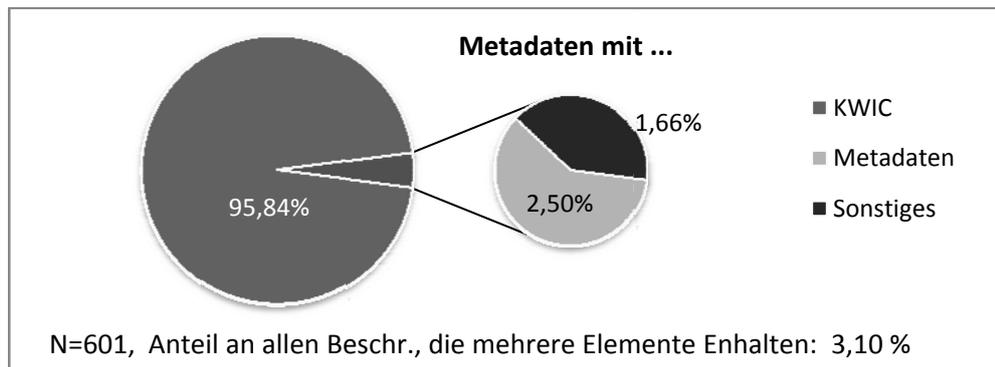


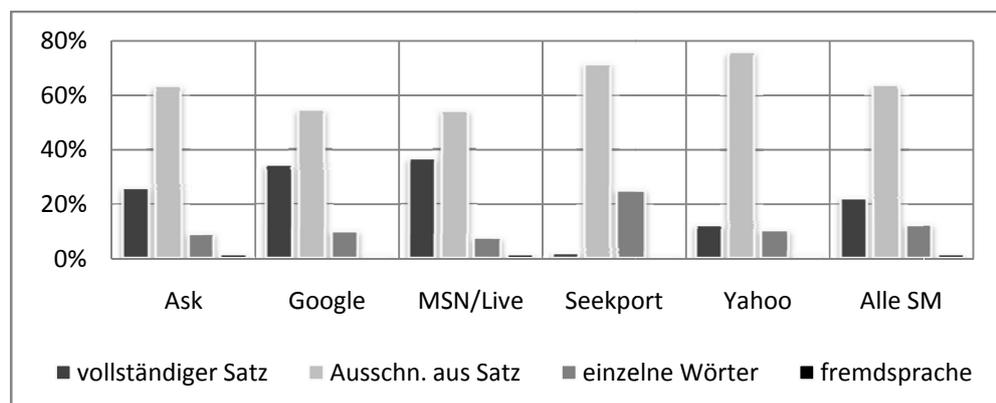
Abbildung 7

Die durchschnittliche Länge der Trefferbeschreibungen beläuft sich zwischen 15,93 Wörtern bei Ask und 39,91 Wörtern bei Seekport. Die Anzahl der Zeichen variieren im Schnitt entsprechend von 119,76 und 303,63 Zeichen. Die maximale Anzahl an Zeichen beträgt bei Ask 176 und bei Seekport 435 Zeichen. Die maximale Anzahl der Wörter erreicht Seekport mit 71 Wörtern. Einen ausführlichen Überblick über die Länge der Beschreibungen liefert die Tabelle 4 und kann darüber hinaus dem Angang A5 entnommen werden.

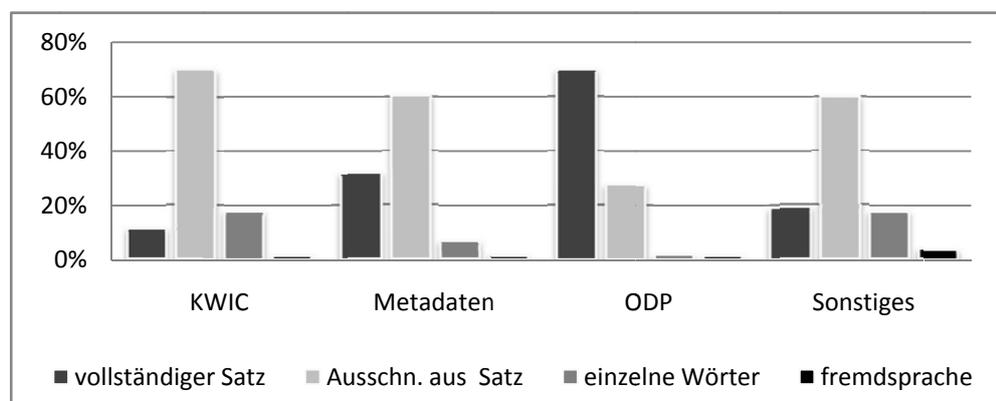
| Suchmaschine | Ask    | Google | MSN    | Seekport | Yahoo  |
|--------------|--------|--------|--------|----------|--------|
| MW Wörter    | 15,93  | 17,89  | 19,01  | 39,32    | 20,30  |
| MW Zeichen   | 119,76 | 135,26 | 143,46 | 303,61   | 149,88 |
| Max Wörter   | 27     | 29     | 42     | 71       | 30     |
| Max Zeichen  | 176    | 197    | 233    | 435      | 229    |

Tabelle 4 Kennzahlen zur Länge der Beschreibungen

Ein weiterer Bestandteil der Untersuchung war die Analyse, wie die Trefferbeschreibungen sprachlich beschaffen sind. Unterschieden wurde dabei (wie in Abschnitt 3.3 erläutert) zwischen vier möglichen Kategorien. Wie man auf Abbildung 8 gut erkennen kann, dominieren in allen Suchmaschinen jene Beschreibungen, die einen oder mehrere Ausschnitte eines Satzes enthalten. Bei den Suchmaschinen Ask, Google und MSN folgen die Beschreibungen, die aus vollständigen Sätzen bestehen. Bei Yahoo liegen die vollständigen Sätze und die Ansammlung einzelner Wörter nahezu gleich auf. Seekport scheint auch hier einen eigenen Weg zu verfolgen und hat neben den Beschreibungen, die Ausschnitte aus Sätzen enthalten, überwiegend Beschreibungen, die eine Zusammenstellung von Wörter enthalten, aus denen sich kein Satz ableiten lässt.



**Abbildung 8: Sprachliche Beschaffenheit je Suchmaschine**



**Abbildung 9: Sprachliche Beschaffenheit je Beschreibungsquelle**

In Bezug auf die Beschreibungsquellen überwiegen auch hier mit Ausnahme jener, die dem ODP entstammen, der Anteil an Ausschnitten aus einem Satz zwischen 60% und 70%, während beim ODP die vollständigen Sätze bei knapp 70% dominieren. Der hohe Anteil an vollständigen Sätzen bei den Beschreibungen, die dem ODP entstammen, ist nachvollziehbar, da es sich beim ODP meist um kurze Beschreibungen der Webseiten handelt. Interessant ist die Tatsache, dass bei den Beschreibungen, die den Metadaten entstammen, der Anteil an Ausschnitten aus einem Satz relativ hoch ist. Dies lässt zwei mögliche Schlüsse zu: Einerseits sind die Angaben in den Metadaten zu lang, um diese vollständig zu übernehmen, oder die Metadaten bzw. der Description Tag enthält häufig unvollständige Sätze.

Ein weiterer Bestandteil der Untersuchung zur Struktur der Trefferbeschreibungen ist die Frage, in wie vielen Treffern die Suchargumente auch in der Beschreibung und wie häufig mehrere Treffer identische Trefferbeschreibungen enthalten sind. Betrachtet man nur jene 4912 Treffer, die auch einen Beschreibungstext enthalten, so enthalten lediglich 2 Treffer nicht die Suchbegriffe. Berücksichtigt man jedoch auch jene, die keine Trefferbeschreibung enthalten und ermittelt jene Treffer, in denen weder in der Trefferbeschreibung noch im Titel des Treffers die Suchbegriffe enthalten, so kommt man auf einen Treffer. Ein Grund für die hohen Quote der Treffer, in denen die Suchbegriffe im Titel oder in der Beschreibung auftauchen, kann man in den relativ kurzen Suchanfragen sehen, da diese ausschließlich aus Anfragen bestehen, die maximal 2 Suchwörter enthalten.

Die Anzahl der identischen Trefferbeschreibungen fallen bei den Suchmaschinen sehr unterschiedlich aus. Während bei Google lediglich 2 Treffer deckungsgleich sind, enthält MSN insgesamt 57 Beschreibungen die in mindestens 2 Treffer enthalten sind. Der Spitzenwert beträgt 9 identische Trefferbeschreibungen auf einer Trefferseite Berücksichtigt man nur jene Treffer, in denen sowohl der Titel als auch die Trefferbeschreibung identisch sind, so erhält man geringere Zahlen, die von keinem Treffer bei Yahoo, bis zu 39 Fällen bei MSN reichen. Den einzelnen Beschreibungsquellen zugeordnet, enthält KWIC insgesamt 39 identische Beschreibungen so-

wie 19 Fälle in denen sowohl der Titel als auch die Beschreibung übereinstimmen. Die Metadaten enthalten dagegen Insgesamt 62 Fälle und davon 33 Fälle in denen der Titel und Beschreibungen übereinstimmen. Den unter Sonstiges erfassten Beschreibungen sind 2 Fälle zugeordnet. Eine detaillierte Übersicht der identischen Treffer ist den Tabelle 5 Tabelle 6 zu entnehmen.

| Anzahl | Ask  | Google | MSN/Live | Seekport | Yahoo | Alle SM |
|--------|------|--------|----------|----------|-------|---------|
| 2      | 17/5 | 1/1    | 43/31    | 10/8     | 11/0  | 82/45   |
| 3      |      |        | 6/5      | 1/1      |       | 10/6    |
| 4      |      |        | 3/1      |          | 1/0   | 4/1     |
| 5      |      |        | 1/0      |          |       | 1/0     |
| 6      | 2/0  |        |          |          |       | 2/0     |
| 7      |      |        | 2/0      |          |       | 2/0     |
| 8      |      |        | 1/1      |          |       | 1/1     |
| 9      |      |        | 1/1      |          |       | 1/1     |
| Summe  | 19/5 | 1/1    | 57/39    | 11/9     | 12/0  | 103/54  |

**Tabelle 5 Anzahl der identischen Trefferbeschreibungen je Suchmaschine**

| Anzahl | KWIC  | Metadaten | Sonstiges | Summe  |
|--------|-------|-----------|-----------|--------|
| 2      | 34/17 | 46/26     | 2/0       | 82/45  |
| 3      | 4/1   | 6/5       |           | 10/6   |
| 4      |       | 4/1       |           | 4/1    |
| 5      |       | 1/0       |           | 1/0    |
| 6      |       | 2/0       |           | 2/0    |
| 7      |       | 2/0       |           | 2/0    |
| 8      |       | 1/1       |           | 1/1    |
| 9      | 1/1   |           |           | 1/1    |
| Summe  | 39/19 | 62/33     | 2/0       | 103/54 |

**Tabelle 6 Anzahl der identischen Trefferbeschreibungen je Beschreibungsquelle**

Dabei gibt die erste Ziffer die Anzahl der insgesamt identischen Beschreibungen an und die zweite Ziffer gibt die Anzahl jener Beschreibungen, in denen sowohl der Titel als auch die Beschreibung deckungsgleich ist, wieder. Der Grund in der hohen Anzahl an identischen Beschreibungen, die den Metadaten entstammen, mag darin liegen, dass häufig bei Seiten, die von einem Content Management System generiert werden, für alle Unterseiten der ein uns selbe Description-Tag im Kopfbereich vergeben wird. Bei jenen identischen Beschreibungen, die aus dem Inhalt des Textes stammen, kann

es einerseits daran liegen, dass ein globales, auf allen Seiten einer Webseite vorhandenes Inhaltselement zurückgegriffen wird und zum anderen, dass die möglicherweise Suchmaschine Schwierigkeiten hat echte Dubletten effektiv zu erkennen.

### **3.5.2 Untersuchung der Wikipedia-Treffer**

Bereits im Vorfeld der Auswertung der Trefferbeschreibungen stellte sich die Frage, wie sich die Suchmaschinen bei Webseiten verhalten, deren formale Struktur weitgehend bekannt ist, und somit die inhaltsbeschreibenden Elemente stets im selben Bereich befinden. Es ist anzunehmen, dass eine Suchmaschine bei der Beschreibung von der relativ homogenen Struktur einer Sammlung von Dokumenten profitieren sollte, da die Suchmaschine stets die gleichen Bereiche in denen beispielsweise ein Abstract befindet, für die Trefferbeschreibung verwenden kann.

Ein solcher Fall ist bei den Einträgen der Web-Enzyklopädie Wikipedia weitgehend gegeben. In den Empfehlungen für gute Artikel heißt es zu Begriffsdefinition und Einleitung: „Sie sollten für sich genommen Lemma und den Begriff in seiner Grundbedeutung erklären.“ (Wikimedia Foundation 2008). Damit sind diese ein geeigneter Kandidat für Beschreibung eines Treffers, der auf einen Wikipedia-Eintrag verweist. Ein weiterer Punkt, der für die Untersuchung der Wikipedia-Treffer spricht ist die Tatsache, dass bei allen Treffern die Beschreibung aus dem Text der Seite selbst stammt. Weiterhin lässt es sich dadurch gut nachvollziehen, aus welchem Bereich einer Seite die Suchmaschine, zur Beschreibung Eintrags zurückgreift.

Da sich bei einer Voruntersuchung gezeigt hat, dass die Suchmaschinen unterschiedlich bei Wikipedia-Treffern vorgehen, wurden alle Treffer, die auf Wikipedia verweisen untersucht, wo genau sich der Text, der für die Trefferbeschreibung verwendet wird, befindet. Unterschieden wurde dabei zwischen folgenden Lokalisierungen:

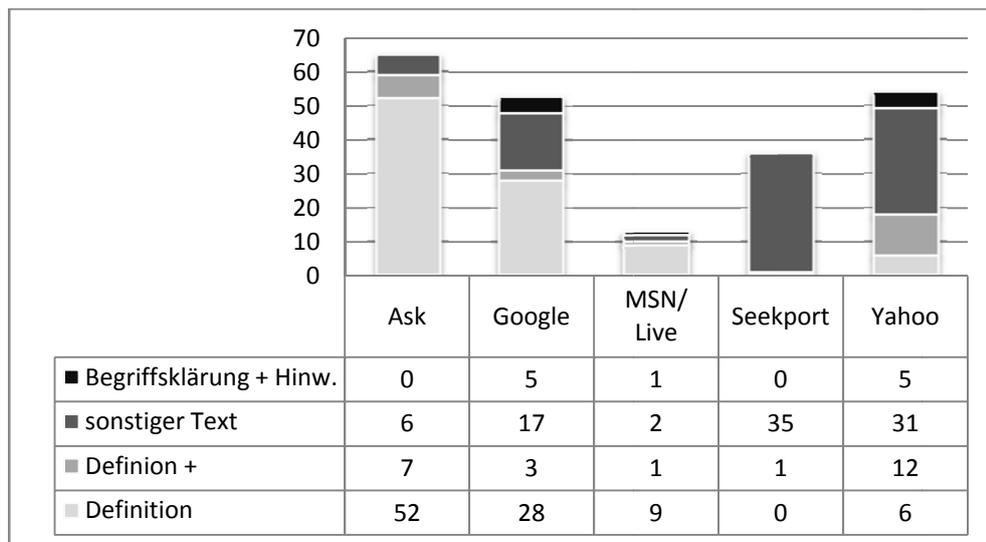
- Einleitender Satz/Definition
- Einleitender Satz/Definition, sowie weiterer Text aus dem Eintrag
- Text, der nicht den Artikel einleitet

- Hinweisseiten und Texte, dazu gehören Begriffsklärungsseiten und Hinweise zu mehrdeutigen Begriffen

Die eindeutige Identifizierung des Seitenbereichs, aus dem die Beschreibung ist (über die Versionsgeschichte, mit dessen Hilfe man bei Wikipedia alle Änderungen nachverfolgen kann), ist auch dann nachvollziehbar, wenn der Eintrag inzwischen geändert wurde.

Von den insgesamt 223 Treffern, die auf Seiten der Domain de.wikipedia.org verweisen, wurden 2 Treffer entfernt, weil diese in einem Fall keine Artikelseite darstellte und im anderen Fall eine Beschreibung enthielt, die aus dem ODP stammt. Die verbleibenden 221 Treffer verteilen sich auf insgesamt 72 Suchanfragen. Auf die Suchmaschine Ask entfallen mit 65 Treffern in 62 verschiedenen Suchanfragen die meisten Wikipedia-Treffer, wohingegen MSN mit 13 Treffern in 12 Suchanfragen die wenigsten Treffer enthält (s. Anlage A10).

Bezogen auf die Beschreibungsquelle wurden insgesamt 119 Treffer identifiziert, die die Definition oder den einleitenden Satz enthalten. Davon entfallen knapp 43% auf Beschreibungen, die nur die Definition bzw. den einleitenden Abschnitt enthalten, sowie etwas unter 11% diejenigen, welche die Definition und weiteren Textauszug aus dem Artikel beinhalten. Bei 91 Treffern wurden andere Textbereiche des Artikels lokalisiert. Hinweistexte und Begriffserklärungen, die den Nutzer über die mehrdeutige Verwendung eines Begriffs aufklären, sind in 11 Treffern vorzufinden. Die Verteilung der Anteile auf einzelnen Suchmaschinen kann der Abbildung 10 entnommen. Diese verdeutlicht die unterschiedlichen „Strategien“ der Suchanbieter. Während Ask nahezu alle Trefferbeschreibungen der Definition bzw. dem einleitenden Abschnitt entnimmt, scheint Seekport fast alle Beschreibungen aus anderen Textabschnitten zu entnehmen. Google und Yahoo entnehmen die Beschreibungen in jeweils umgekehrten Proportionsverhältnissen den Definitionen bzw. den sonstigen Textbereichen (Google: ca. 58%:32% Yahoo ca. 33%:57%).



**Abbildung 10 Lokalisierung der Beschreibungsquellen auf Wikipedia**

### 3.5.3 Zusammenfassung der Strukturanalyse der Trefferbeschreibungen

Bezugnehmend auf die erste Fragestellung der Untersuchung „Wie häufig werden die verschiedenen Beschreibungsquellen von den Suchmaschinen genutzt?“ kann festgestellt werden, dass die die Ausschnitte aus dem Text und die Metadaten dominieren. Bei einigen Suchmaschinen liegt das Gewicht etwas stärker auf dem KWIC-Ansatz, anderen scheinen wiederum Metadaten zu bevorzugen. Der Anteil der Beschreibungen, die aus dem ODP stammen, liegt bei allen Suchmaschinen unter 10%. Bei Seekport fällt auf, dass diese statt einem Mix aus verschiedenen Beschreibungsquellen fast ausschließlich die KWIC-Methode als Beschreibungsquelle nutzt. Interessant ist die Tatsache, dass nahezu alle Treffer die Suchworte enthalten. Ein Grund mag einerseits darin liegen, dass die Suchanfragen, die maximal zwei Wörter enthalten, relativ kurz sind und die Chancen somit relativ hoch sind, dass der Beschreibungstext die Suchbegriffe enthält. Denkbar wäre andererseits aber auch, dass die Suchmaschinen von den verschiedenen zur Verfügung stehenden Beschreibungsquellen jene bevorzugt, die die entsprechen-

den Begriffe enthalten. Im Zusammenhang mit den identischen Trefferbeschreibungen lässt sich feststellen, dass die Suchmaschinen unterschiedlich effektiv solche „Beschreibungs-Dubletten“ erkennen.

Die zweite Untersuchungsfrage beschäftigte sich mit den charakteristischen Merkmalen einer Trefferbeschreibung, wie zum Beispiel deren Länge oder sprachliche Beschaffenheit.

Die Untersuchung der Länge der Beschreibungen ergab bei allen Suchmaschinen (bis auf Seekport) ähnliche Ergebnisse. Die vier Suchmaschinen kommen zusammen auf einen Durchschnitt von 18,28 Wörtern bzw. 137 Zeichen, was einer Zeilenlänge von ca. 1,5 Zeilen entspricht, wenn man von 90-100 Zeichen pro Zeile ausgeht. Die Beschreibungen von Seekport fallen mit durchschnittlich 303 Zeichen mehr als doppelt so lang aus, als der Durchschnitt der anderen vier Suchmaschinen.

Es konnte festgestellt werden, dass in allen Suchmaschinen mehr als ein Ausschnitt aus der Seite bzw. mehrere Beschreibungsquellen verwendet werden. Allerdings unterscheiden sich die Suchmaschinen in der Nutzung dieser mehrteiligen Beschreibungselemente. Während Ask, Google und MSN ihren Schwerpunkt auf einteilige Beschreibungen legen, dominieren bei Yahoo und Seekport Beschreibungen, die mindestens 2 Elemente enthalten.

Eine Folgeuntersuchung ergab, dass in dem zweiten Beschreibungselement die Ausschnitte aus dem Text deutlich dominieren. Bezogen auf alle Trefferbeschreibungen (also auch solche, die nur ein Element enthalten) beträgt der Anteil jener Treffer, deren Beschreibungen im ersten oder zweiten Element mindestens einen Ausschnitt aus dem Inhalt des Dokuments enthalten, über 60%

Ein weiterer Bestandteil war die Analyse der sprachlichen Beschaffenheit der Trefferbeschreibungen. Der größte Anteil der Treffer enthielt einen oder mehrere Ausschnitte aus einem Satz (>50%). Der Anteil der vollständigen Sätze liegt bei den Suchmaschinen Google und MSN über 30% während dieser bei den verbleibenden Suchmaschinen zwischen 2% und 20% lag. Bei den Beschreibungsquellen konnte bis auf ODP eine Dominanz der un-

vollständigen Sätze festgestellt werden, während das ODP überwiegend vollständige Sätze enthielt, was auch zu erwarten war, da diese nach den Editier-Richtlinien informativ, prägnant und kurz gehalten werden sollten (Open Directory Project 2004) und somit in der Regel kurz genug ausfallen, um vollständig als Trefferbeschreibung übernommen zu werden.

Die Analyse der Wikipedia-Treffer ergab, dass die Suchmaschinen unterschiedliche Seitenbereiche für die Beschreibung nutzen. Es wurde angenommen, dass die Definition bzw. der einleitende Satz der geeignetste Bereich für eine Beschreibung eines Wikipedia Eintrags ist. Einzig Ask scheint diese Strategie weitgehend bis auf wenige Ausnahmen zu verfolgen. Bei den verbleibenden Suchmaschinen überwiegen entweder jene Treffer, die die Definition bzw. die Einleitung enthalten oder jene die einen Ausschnitt aus einem anderen Bereich der Seite enthalten. Allerdings fällt hier Seekport etwas heraus, da der Anteil der Beschreibungen, die Ausschnitte enthalten, die nicht aus der Anfang der Seite stammen, nahezu gänzlich überwiegt.

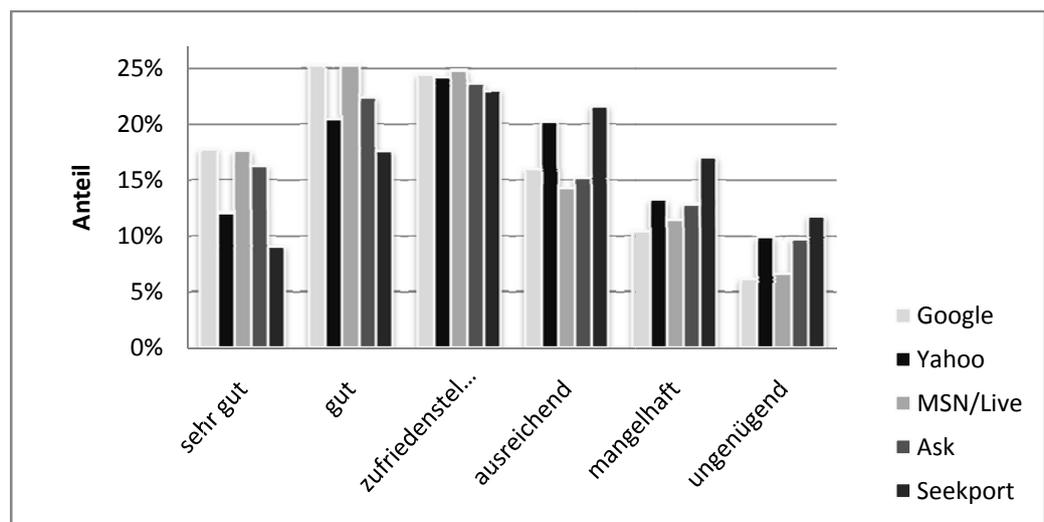
#### **3.5.4 Bewertung der Treffer mit Fragebögen**

Ziel dieser Bewertung der Trefferbeschreibungen mit Hilfe eines Fragebogens war es Erkenntnisse zu erhalten, inwiefern sich Unterschiede in Bezug auf die Charakteristika einer Beschreibung ergeben. Ein weiterer Bestandteil der Befragung war die Erhebung der bevorzugten Suchmaschine, sowie die Angaben der Befragten, wie eine ideale Trefferbeschreibung beschaffen sein sollte.

Insgesamt wurden 4733 Treffer von 97 Personen bewertet. Das entspricht einem Anteil von 94,15% an der Gesamtanzahl der Treffer, die bei der Analyse der Trefferbeschreibungen ausgewertet wurden (s. Abschnitt 3.5.1).

Die Anzahl der Missings beträgt 26 Treffer, die sich daraus ergaben, dass entweder versehentlich eine Seite überblättert wurde, oder Unklarheit bestand, wie Treffer, die nur einen Titel (also keinen Beschreibungstext) enthielten, zu bewerten sind.

Um die Bewertungen der einzelnen Charakteristika beschreiben und vergleichen zu können, wurden die Anteile der Notenwerte je Merkmalgruppe ermittelt. Die Verwendung der Anteile statt der absoluten Werte ist notwendig, damit trotz der unterschiedlichen Anteile an der Gesamtzahl der Treffer die Anteile der Notenwerte je Merkmalgruppe vergleichbar sind. Als eine weitere Kennzahl wurde das arithmetische Mittel der vergebenen Noten je Merkmalgruppe ermittelt.<sup>16</sup>



**Abbildung 11** Anteile der Noten je Suchmaschine

Betrachtet man die Bewertungen der Trefferbeschreibungen im Zusammenhang mit den einzelnen Suchmaschinen, so ergibt sich insgesamt ein relativ vergleichbarer Anteilsverlauf (s. Abbildung 11). Der Chi-Quadrat-Test ergab, dass die Suchmaschinen Google und MSN in ihren Anteilen kaum unterscheiden ( $p > 0,90$ ), sowie die Suchmaschinen Ask und MSN geringfügig

<sup>16</sup> Da es sich bei Noten um eine Ordinalskala handelt, wurden die Notenwerte in ein Punktesystem übertragen, um so die entsprechenden Mittelwerte zu erhalten. Die Verwendung der Mittelwerte der Noten erscheint dem Verfasser entgegen der Auffassung der Statistik, dass erst ab metrischen Skalen die Möglichkeit besteht, Mittelwerte zu ermitteln, legitim zu sein, da die Durchschnittsnoten sehr verbreitet und allgemein verständlich sind und damit interpretierbar sind.

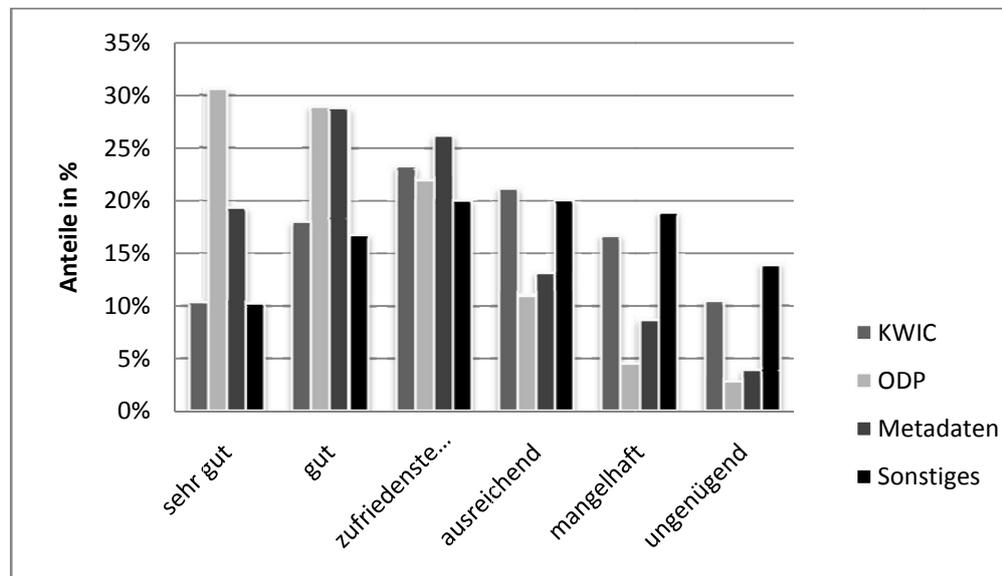
gig unterscheiden ( $p = 0,112$ ), während die anderen Suchmaschinen untereinander Signifikant unterscheiden ( $p < 0,03$ ).

Betrachtet man statt den Anteilen an den Notenwerten die Mittelwerte, so kann man feststellen, dass alle Suchmaschinen Google, MSN und Ask relativ nah beieinander liegen, während Yahoo und Seekport schlechtere Durchschnittswerte haben.

|          | Mittelwert | N     |
|----------|------------|-------|
| Google   | 2,9464     | 970   |
| Yahoo    | 3,3184     | 939   |
| MSN/Live | 2,9673     | 918   |
| Ask      | 3,1522     | 966   |
| Seekport | 3,5543     | 940   |
| Alle SM  | 3,1870     | 4.733 |

**Tabelle 7 Mittelwerte der Bewertungen je Suchmaschine**

Betrachtet man dagegen, wie die Bewertungen der einzelnen Beschreibungstypen ausfallen, so kann man feststellen, dass die verschiedenen Beschreibungsquellen unterschiedliche Notenschwerpunkte haben. Hierbei wurden nur jene Treffer berücksichtigt, die einen Beschreibungstext enthalten.



**Abbildung 12 Anteile der Bewertungen je Beschreibungsquelle**

Wie auf Abbildung 12 gut zu erkennen ist, liegt der Schwerpunkt der Beschreibungen, die dem ODP entstammen, in dem vorderen Notenbereich und nimmt zunehmend ab, je schlechter der Notenwert ist. Bei den Beschreibungen, die den Metadaten entnommen sind, liegt der Schwerpunkt ebenfalls im vorderen Notenbereich, jedoch mit einem geringeren Anteil der Note sehr gut. Die Beschreibungen, die dem KWIC und keiner Quelle zugeordnet werden konnten, haben ungefähr gleich hohe Notenanteile, die in allen Notenwerten bei mindestens 10% liegen. Die Mittelwerte geben den Trend der Anteile noch deutlicher wieder: So führt ODP, gefolgt von den Metadaten, während KWIC im Vergleich zum ODP etwas mehr als einen Notenwert schlechter ausfällt, gefolgt von den nicht zugeordneten Beschreibungen. Allerdings können diese ermittelten Anteile und Durchschnittswerte nur mit einer gewissen Einschränkung für die einzelnen Beschreibungsquellen als repräsentativ gesehen werden, da die absoluten Anteile sehr unterschiedlich ausfallen. Andererseits spiegeln diese wiederum repräsentativ die derzeitige Verwendung der Beschreibungsquellen seitens der Suchmaschinen wieder.

| Beschreibungstyp | Mittelwert   | N            |
|------------------|--------------|--------------|
| KWIC             | 3,474        | 2.406        |
| ODP              | 2,384        | 245          |
| Metadaten        | 2,749        | 1.739        |
| Sonstiges        | 3,623        | 244          |
| <b>Insgesamt</b> | <b>3,151</b> | <b>4.634</b> |

**Tabelle 8 Mittelwerte der Beschreibungsquellen**

|                           | Mittelwert   | N            |
|---------------------------|--------------|--------------|
| Vollständiger Satz        | 2,647        | 1.043        |
| Ausschnitt aus einem Satz | 3,149        | 2.952        |
| Einzelne Wörter           | 3,974        | 579          |
| <b>Insgesamt</b>          | <b>3,151</b> | <b>4.634</b> |
| TB in Fremdsprache        | 4,100        | 60           |

**Tabelle 9 Mittelwerte sprachliche Beschaffenheit der TB**

Es wurde angenommen, dass sprachlich vollständigeren Beschreibungen bessere Bewertungen erhalten. Die Bewertungen der sprachlichen Beschaffenheit der Trefferbeschreibungen fallen insgesamt dementsprechend erwartungsgemäß aus. So entfällt auf Beschreibungen, die vollständige Sätze enthalten, der höchste Anteil der mit gut und besser bewerteten Noten, sowie ein Anteil an den mangelhaft und ungenügend von unter 10%. Bei den unvollständigen Sätzen liegt der Schwerpunkt im mittleren Notenbereich, während die Beschreibungen, die aus einzelnen Wörtern bestehen, ihren Schwerpunkt eher in den negativen Noten haben (s. Abbildung 13). Auch anhand der Mittelwerte lässt sich diese Rangfolge von vollständigen Sätzen über Ausschnitte aus einem Satz und die einzelnen Wörter als am schlechtesten abschneidenden Sprachtyp feststellen (Tabelle 9). Auf die Berücksichtigung der Trefferbeschreibungen, die in einer Fremdsprache vorliegen, wurde verzichtet, da deren Anteil an der Gesamtsumme mit 2% sehr gering ausfällt. Die Untersuchung des Zusammenhangs mittels Chi-Quadrat-Test in Verbindung mit dem Cramerschen Assoziationsmaß, das die Stärke des Zusammenhangs bei einer Skala von 0 - 1 misst, ergab sowohl bei den Beschreibungsquellen, als auch bei den Sprachtypen ein schwacher Zusammenhang bei einem V von unter 0,190.

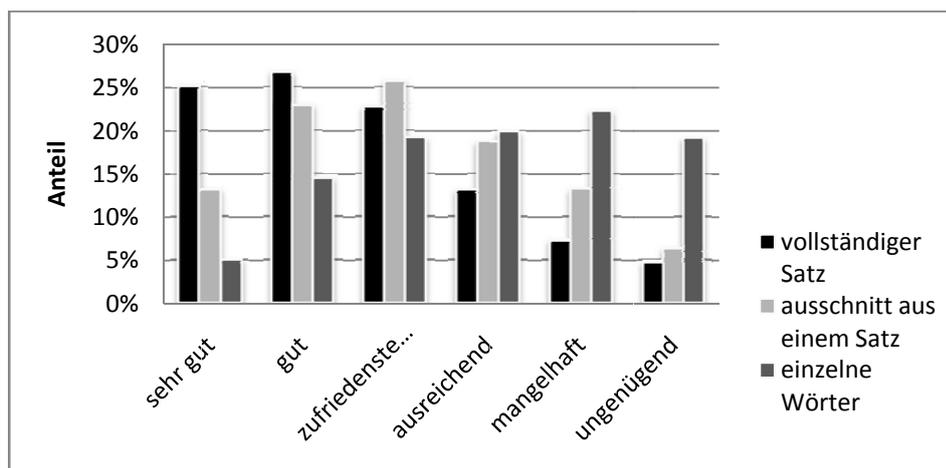
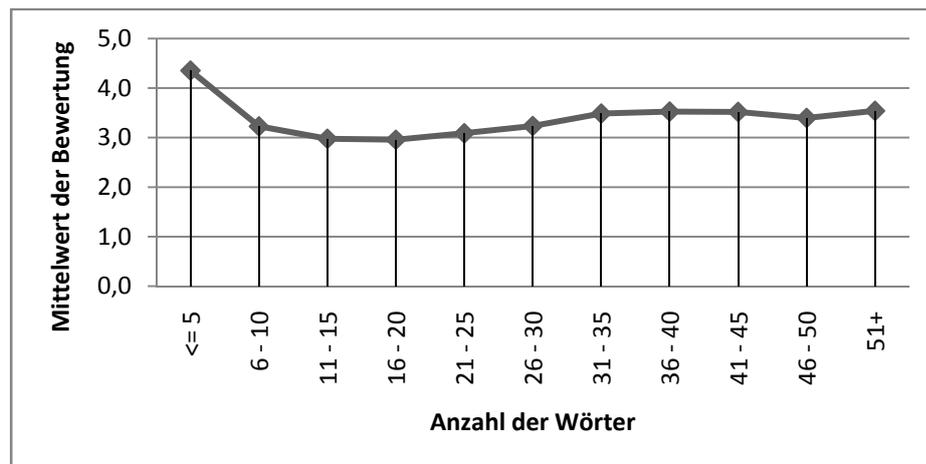


Abbildung 13 Anteile der Noten je Sprachtyp

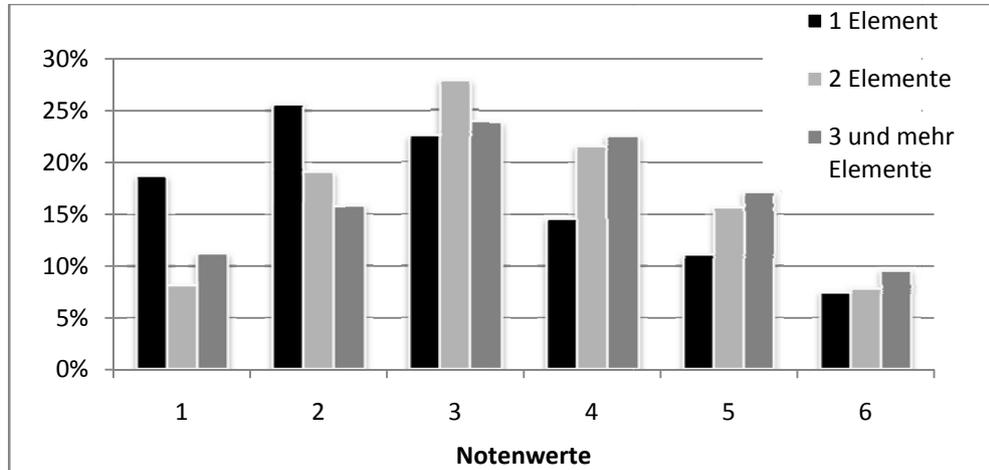
Betrachtet man die Mittelwerte der Bewertungen im Zusammenhang mit der Länge der Trefferbeschreibungen, so kann man feststellen, dass bei sehr kurzen Beschreibungen, die bis zu 10 Wörter enthalten, der Mittelwert etwas schlechter ist, während ab der Klasse mit mindestens 11 Wörtern der Mittelwert am besten ist und sich von da an sich leicht, aber kontinuierlich verschlechtert, je mehr Wörter in der Trefferbeschreibung enthalten sind (s. Abbildung 14 sowie Anhang A11).



**Abbildung 14 Mittelwerte und Länge der TB (in Klassen)**

Um die Anzahl Elemente als Einflussfaktor der Bewertungen besser betrachten zu können, wurde die Anzahl der Elemente in drei Klassen unterteilt: Jene die nur ein Element enthalten, jene die 2 Elemente enthalten sowie eine Klasse mit allen Beschreibungen, die drei oder mehr Elemente enthalten. Da keine Suchmaschine außer Seekport einen nennenswerten Anteil an Beschreibungen, die mehr als 3 Elemente enthalten, erschien die Bündelung der Beschreibungen, die mindestens 3 Elemente enthalten, als sinnvoll. Vergleicht man die Bewertungen dieser drei „Elementeklassen“, so kann man feststellen, dass Beschreibungen die nur ein Element enthalten bessere Bewertungen erhalten haben, als jene Beschreibungen, die mehr als ein Element enthalten. Weiterhin erhöht sich der Anteil der negativen Noten in der Klasse, die aus mehr als drei Elementen bestehen. Noch deutlicher

wird der Unterschied, wenn man die Beschreibungen, die nur aus einem Element bestehen, direkt mit jenen vergleicht, in denen mehrere Elemente vorkommen.



**Abbildung 15** Bewertungen je Anzahl der Elemente

Die Frage nach der bevorzugten Suchmaschine, nach der die Teilnehmer im Anschluss an der Bewertung befragt wurden, ergab dass 95 der 97 Befragten Google als die bevorzugte Suchmaschine angaben. Ein Befragter nannte Yahoo und eine weitere Person verzichtete auf die Angabe der Suchmaschine. Insgesamt 7 Personen nannten zusätzlich weitere Suchmaschinen. Dabei wurde 6 mal Yahoo und Ask 2 mal genannt. Unter Sonstiges, also jene Suchmaschinen, die nicht in der Untersuchung selbst berücksichtigt wurden, wurde die Suchmaschine Metager, sowie Wikipedia genannt. Allerdings ist Wikipedia nicht unmittelbar als eine Suchmaschine zu betrachten. Im Zusammenhang mit diesen bevorzugten Suchmaschinen stellt sich die Frage, wie häufig die vom Befragten am besten bewerteten Suchmaschine mit der vom Befragten bevorzugten Suchmaschine übereinstimmt. Dazu wurden die Mittelwerte zu jedem Fragebogen-Set, das gleichzeitig einem Befragten entspricht, ermittelt und anschließend verglichen, ob eine Übereinstimmung mit einer der bevorzugten Suchmaschinen vorliegt. Bei einem Drittel der 96 Fälle, in denen eine bevorzugte Suchmaschinen angegeben wurde, lag eine

Übereinstimmung zwischen der am besten bewerteten Suchmaschine und der bevorzugten Suchmaschine vor. Die Abweichungen und welche Suchmaschine statt der bevorzugten Suchmaschine am besten bewertet wurde, kann der Tabelle 10 entnommen werden. Die Zeilen geben die bevorzugten Suchmaschinen wieder und den Spalten sind die am besten bewerteten Suchmaschinen je Befragter zu entnehmen. Die Anteile der am besten bewerteten Suchmaschinen je Befragter ist der untersten Zeile zu entnehmen und bestätigt die Rangfolge der Suchmaschinen, die weiter oben in Bezug auf alle Treffer ermittelt wurde.

| Bevorzugte SM |       | Am besten bewertete Suchmaschine |               |               |              |               | Alle SM        |
|---------------|-------|----------------------------------|---------------|---------------|--------------|---------------|----------------|
|               |       | Ask                              | Google        | MSN/Live      | Seekport     | Yahoo         |                |
| Google        |       | 16                               | 30            | 25            | 6            | 11            | 88             |
|               | Ask   |                                  | 1             |               |              |               | 1              |
|               | Yahoo |                                  | 1             | 4             |              |               | 5              |
|               | Yahoo | Ask                              |               |               | 1            |               | 1              |
| Yahoo         |       | 1                                |               |               |              |               | 1              |
| Keine Angabe  |       |                                  |               | 1             |              |               | 1              |
| <b>Summe</b>  |       | <b>17</b>                        | <b>32</b>     | <b>31</b>     | <b>6</b>     | <b>11</b>     | <b>97</b>      |
|               |       | <b>17,53%</b>                    | <b>32,99%</b> | <b>31,96%</b> | <b>6,19%</b> | <b>11,34%</b> | <b>100,00%</b> |

**Tabelle 10 Bevorzugte und am besten bewertete Suchmaschinen**

Die Frage was den Befragten bei einer Trefferbeschreibung besonders wichtig sei, ergab, dass von den 96 Personen 40 die Beschreibung in ganzen Sätzen als wichtigstes Merkmal genannt haben. Die Ausschnitte der Textstellen, in denen die Suchwörter vorkommen, wurden 34 mal genannt. Einen Ausschnitt aus dem Inhalt der Seite, wobei nicht näher definiert wurde, wie dieser beschaffen ist, wurde 19 mal genannt und die Auszüge einzelner Stichwörter bzw. Kategorien erhielt 3 Nennungen. Obwohl nur eine Angabe gewünscht war, kreuzten 10 Personen mehrere Antworten an. Von jenen 40 Personen, die den vollständigen Satz angegeben haben, nannten 4 Personen zusätzlich auch den Ausschnitt aus den Textstellen, sowie eine Person die Kategorien bzw. Stichwörter. Von den 34 Personen, die Ausschnitte aus

einem Satz als besonders wichtig angegeben haben, haben 5 Personen auch den Ausschnitt der Textstellen, in denen die Suchwörter vorkommen, genannt. Damit wurde der Ausschnitt der Textstellen, in denen der Suchbegriff vorkommt, insgesamt 43 mal genannt. Addiert man die spezifischere Aussage, in der zusätzlich zum Auszug aus dem Text der Bereich festgelegt wird, in denen die Suchbegriffe vorkommen, zu jener Aussage, die einen nicht näher definierten Ausschnitt aus dem Inhalt einer Seite nennt, so kommt man auf einen Anteil von über 60%.

Befragt nach der idealen Länge einer Trefferbeschreibung, nannten von den 96 Personen 48% drei Zeilen, 28% zwei Zeilen sowie 16 % vier Zeilen und eine Zeile bzw. fünf Zeilen und mehr Zeilen wurden von jeweils 4 Prozent als ideale Länge angesehen. Im Durchschnitt ergibt sich daraus eine Länge von gerundet drei Zeilen.

### **3.5.5 Zusammenfassung der Befragung zur Bewertung der Trefferbeschreibungen**

Im Folgenden sollen die Ergebnisse der Bewertung in Verbindung mit der dritten Kernfrage der Untersuchung „Wie werden die Beschreibungsquellen von den Nutzern angenommen?“ betrachtet werden.

Die Betrachtung der Suchmaschinen ergab sowohl bezogen auf die Trefferbeschreibungen, als auch im Zusammenhang mit den Befragten ähnliche Ergebnisse. Google und MSN führen nahezu gleichauf, gefolgt von Ask, während Yahoo und Seekport etwas abgeschlagen sind. Allerdings sind die Unterschiede zwischen den Suchmaschinen bei den Anteilen, bei wie vielen Befragten welche Suchmaschine am besten abgeschnitten hat, größer, jedoch liegen auch hier Google und MSN nahezu gleichauf. Die Untersuchung der Übereinstimmungen zwischen der bevorzugten Suchmaschine, bei der über 98% Google genannt wurde, ergab, dass bei 32% eine Übereinstimmung vorlag, während sich bei 60% der Befragten sich die bevorzugte und tatsächlich am besten bewertete Suchmaschine unterschieden.

Die Bewertungen im Blickwinkel der Beschreibungsquellen ergab, dass die Beschreibungen, die dem ODP entstammen, führen, mit leichtem Abstand gefolgt von den Metadaten. KWIC und die keiner Quelle zugeordneten Trefferbeschreibungen folgen mit einem höheren Anteil an negativen Noten und somit auch schlechteren Durchschnittswerten, die mehr als einen Notenwert hinter jenem Durchschnittswert von ODP liegen.

Am deutlichsten sind Unterschiede bei der sprachlichen Beschaffenheit der Trefferbeschreibungen festzustellen. So führen die vollständigen Sätze, die in den Noten gut und sehr gut dominieren, vor den Ausschnitten aus einem Satz. Die Ansammlungen von einzelnen Wörtern liegen eher im negativen Notenbereich und erhalten somit auch eine wesentlich schlechtere Durchschnittsbewertung.

Befragt was dem Befragten bei der Trefferbeschreibung besonders wichtig sei, gaben jeweils über 40% an, das seien Beschreibungen in vollständigen Sätzen, als auch Ausschnitte aus dem Inhalt der Seite, in denen die Suchbegriffe vorkommen. Betrachtet man jene, die einen Ausschnitt aus dem Inhalt einer Seite, unabhängig davon, ob die Suchbegriffe vorkommen, genannt haben, so kommt man auf einen Anteil von knapp zwei Drittel. Dass sowohl der Ausschnitt der Textstellen in denen die Suchbegriffe vorkommen, als auch die Trefferbeschreibungen in vollständigen Sätzen von über 40% der Befragten genannt wurden, zeigt einerseits, dass ein Zusammenhang mit der Suchanfrage erkennbar sein sollte und zum anderen, dass eine sprachlich verständliche, vollständige Beschreibung gewünscht ist. Andererseits zeigt gerade die häufige Nennung der Ausschnitte eines Satzes in denen die Suchbegriffe vorkommen, wie stark die Gewohnheit der Suchmaschinen-nutzung und andererseits die bisherige Praxis der großen Suchmaschinen, die Suchbegriffe in den Treffern hervorzuheben, bei den Befragten verankert ist.

Im Zusammenhang mit der Länge der Trefferbeschreibungen konnte eine Auswirkung auf die Bewertungen festgestellt werden. Sehr kurze Trefferbeschreibungen erhielten schlechtere Bewertungen und ab der Klassen, die 10-15 Wörter enthalten, ist eine leichte Verschlechterung der Durchschnittsbe-

wertungen festzustellen, je länger die Bewertungen ausfallen. Befragt nach der idealen Länge einer Trefferbeschreibungen, nannten knapp die Hälfte der Befragten drei Zeilen, gefolgt von zweizeiligen Beschreibungen.

Im Zusammenhang mit der vierten und letzten Untersuchungsfrage, die nach den möglichen Einflussfaktoren der Bewertung einer Trefferbeschreibung fragte, kann angenommen werden, dass mehrere Faktoren einen Einfluss auf die Beurteilung eines Treffers haben. Als Faktoren konnten im Rahmen dieser Untersuchung einerseits die sprachliche Beschaffenheit bzw. Qualität, sowie die Länge einer Trefferbeschreibung identifiziert werden. Da bei den einzelnen Beschreibungsquellen bestimmte sprachliche Niveaus häufiger vorkommen, beeinflusst die sprachliche Beschaffenheit auch die Beurteilung der Beschreibungsquellen. Selbiges gilt für die Länge der Trefferbeschreibungen.

### **3.6 Diskussion der Untersuchungen**

Die Untersuchung hat gezeigt, dass alle Suchmaschinen, bis auf Seekport, auf einen Mix von verschiedenen Beschreibungsquellen setzen und je nach Suchmaschine mehr oder weniger häufig mehrere Beschreibungselemente verwenden, um ein Dokument zu beschreiben. Die Bewertung der Treffer in Bezug auf die Aussagekraft der Trefferbeschreibung zeigt, dass die Quellen, die nicht dem derzeit am häufigsten verwendeten KWIC-Ansatz folgen, ein höheres Potential bieten, aussagekräftige Beschreibungen zu liefern. Dies trifft insbesondere auf die Beschreibungen zu, die dem ODP entstammen. Die Verwendung der Metadaten, die am zweithäufigsten verwendet werden, scheinen aufgrund der häufig zu langen Beschreibungen an Aussagekraft einzubüßen, da dadurch nur ein Ausschnitt der hinterlegten Beschreibung angezeigt wird. Einer der Schlüssel-Faktoren zu einer aussagekräftigen Beschreibung scheint in der sprachlichen Vollständigkeit zu liegen, wie neben der Bewertung selbst auch die Erhebung, welche Eigenschaften die ideale Trefferbeschreibung haben sollte, gezeigt hat. Jedoch besteht auch ein Bedürfnis, dass der Zusammenhang mit der Suchanfrage in der Trefferbeschreibung nachvollziehbar sein sollte, wie die sehr häufige Nennung des

Ausschnittes des Inhalts, in dem die Suchbegriffe vorkommen, zeigt. Diese Erkenntnisse spiegeln sich auch in einer früheren Studie von Machill u.a., wieder, in der die Beschaffenheit der Beschreibungen der in den Laborexperiment aktivierten Treffer am häufigsten einen sinnvollen Beschreibungstext und am zweithäufigsten einen sinnvollen Zusammenhang mit den Suchwörter enthielten (Machill u.a. 2003, S. 270).

Eine generelle Empfehlung für oder gegen eine bestimmte Suchmaschine kann anhand der vorliegenden Untersuchung nicht gemacht werden. Zwar schneiden die Suchmaschinen Google und MSN etwas besser ab als die anderen, aber das scheint insbesondere daran zu liegen, dass deren Anteil an vollständigen Sätzen höher ist als bei den verbleibenden Suchmaschinen. Ein besseres Abschneiden, einer deutschen Suchmaschine, die auf die deutsche Sprache entwickelt und ausgerichtet ist, konnte nicht festgestellt werden. Im Gegenteil, das Konzept von Seekport, nur Ausschnitte aus dem Inhalt der Seite als Quelle zu verwenden, sowie lange, aus mehreren Elementen bestehenden Trefferbeschreibungen zu verwenden, scheinen sich negativ auf die Aussagekraft einer Trefferbeschreibung auszuwirken.

Da die einzelnen Beschreibungsquellen sehr unterschiedlich hohe Anteile an der Gesamtanzahl der untersuchten Trefferbeschreibungen haben, können die Erkenntnisse, inwiefern bestimmte Beschreibungsquellen geeigneter sind einen Inhalt aussagekräftig zu beschreiben, nur bedingt repräsentativ. Dies wurde jedoch hingenommen, da Beibehaltung der tatsächlich von den Suchmaschinen verwendeten Anteile im Vordergrund stand, um ein realistisches Abbild der derzeitigen Trefferstruktur zu bieten.

Andererseits wäre es für die Untersuchung der Beurteilung vorteilhaft gewesen, wenn ähnlich große Häufigkeiten der Beschreibungsquellen vorhanden gewesen wären.

Aufgrund der Tatsache, dass im Rahmen dieser Untersuchung ein relativ hoher Anteil sehr kurzer Suchanfragen verwendet wurde, bleibt die Frage offen, ob und inwiefern längere Suchanfragen zu anderen Trefferbeschreibungen führen. Denkbar wäre, dass möglicherweise die Häufigkeit der ver-

wendeten Elemente auch bei jenen Suchmaschinen steigt, die ihren Hauptanteil bei jenen Beschreibungen haben, die nur ein Element enthalten.

In Bezug auf die Bewertung wäre sicherlich ein kürzerer Fragebogen und dafür eine höhere Anzahl an Befragten sinnvoll gewesen, da der verwendete Fragebogen mit jeweils 50 Treffern relativ lang war und somit ein erhöhtes Risiko an Ermüdungserscheinungen seitens der Befragten gegeben war. Allerdings hätte eine doppelte Anzahl an Befragten den Rahmen an zeitlichen und organisatorischen Aufwand gesprengt, so dass diese möglichen „Beeinträchtigungen“ durch Ermüdungserscheinungen hingenommen wurden.

Im Rahmen dieser Untersuchung konnte die Entscheidungsgründe im einzelnen, warum bestimmte Beschreibung für geeigneter gehalten werden, nicht erhoben werden, da eine solche Befragung sehr zeitintensiv und mit erheblichen Aufwand verbunden wäre. Jedoch besteht hier Forschungsbedarf, um Erkenntnisse zu erlangen, wie die Beschreibungen genutzt und welche Kriterien dabei eine Rolle spielen. Die Erkenntnisse aus einer solchen Erhebung können auch die qualitative Verbesserung Trefferbeschreibungen fördern.

Offen bleibt nach dieser Untersuchung die Frage, wie ausgeprägt das Angebot an verschiedenen Beschreibungsquellen ist, also wie häufig in den Metadaten ein Description Tag, oder ein Eintrag im ODP zu einer Webseite vorhanden ist. Eine entsprechende Erhebung, in wie vielen Fällen solche Angaben vorhanden sind, wäre steht noch aus.

#### **4 Entwicklung einer idealen Trefferbeschreibung**

Auf Basis der Erkenntnisse der vorhergehenden Untersuchung soll im Folgenden ein Schema zur Generierung von Trefferbeschreibungen entwickelt werden, sowie Optimierungsmöglichkeiten sowohl seitens der Suchmaschinenbetreiber, als auch Webseitenbetreiber aufgezeigt werden.

Die derzeit von den Suchmaschinen verwendeten Beschreibungsquellen haben unterschiedliche Stärken und Schwächen. Der Vorteil des KWIC-

Ansatzes besteht darin, dass die dafür benötigten Informationen in der Regel im Dokument vorhanden sind. Eine Ausnahme stellen jene Inhalte dar, die von den Suchmaschinen nicht erfasst werden können, z.B. Flash-Inhalte etc. Weiterhin ist für den Nutzer durch die Hervorhebung der Suchbegriffe nachvollziehbar, in welchem Kontext die Suchbegriffe auftreten. Andererseits kann es vorkommen, dass aus dem Ausschnitt isoliert für sich alleine Betrachtet der Gesamtkontext des Dokuments nicht erschließbar ist. Nachteilig wirkt häufig der Umstand aus, dass in der Regel nur ein Fragment eines Satzes entnommen wird, so dass die sprachliche Qualität der Beschreibung dadurch eingeschränkt wird.

Die Stärke der beschreibenden Metadaten liegt darin, dass diese idealerweise eine kurze, prägnante Beschreibung des Dokuments enthalten. Jedoch stellen sie ein optionales Element eines Dokuments dar, das zusätzlich erstellt werden muss. Ein gewisses Missbrauchspotential der Metadaten besteht darin, dass die Metadaten Informationen enthalten, die nicht den Inhalt des Dokuments beschreiben, sondern mit dem Motiv erstellt wurden entweder eine bessere Ranking-Position zu erhalten, oder Interessanten eines fremden, populäreren Themengebieten auf die Webseite zu locken. Dies geschah in der Vergangenheit insbesondere bei dem Keyword-Tag, der aufgrund dessen heute für die Suchmaschinen eine untergeordnete Rolle spielt (Erlhofer 2007b, S. 45).

Die Beschreibungen in Webverzeichnissen haben im Grunde ähnliche Eigenschaften, wie die Metadaten, jedoch werden diese von dem Dokument aus betrachtet externen Personen erstellt, so dass eine höhere Neutralität als bei den Metadaten zu erwarten ist. Die Schwäche solcher Verzeichnisse besteht darin, dass es schier unmöglich ist, für jedes Web-Dokument eine Beschreibung zu erstellen. Realistischer ist es für Webseiten und ggf. den Hauptunterseiten solche Beschreibungen zu erwarten.

Im Zusammenspiel mit den verschiedenen Anfragetypen können die Beschreibungsquellen unterschiedlich profitieren. So haben die externen Beschreibungen in den Verzeichnissen, als auch die internen Beschreibungen bei navigationsorientierte Anfragen, ein höheres Potential, den Nutzer mit

einer aussagekräftigen Beschreibung bei der Evaluation der Treffer zu unterstützen. Vergleichbares im ähnlichen Umfang gilt auch für transaktionsorientierte Anfragen, wobei hier möglicherweise in den Metadaten ein höheres Potential liegen kann, da vom Betreiber erstellte Beschreibungen das Angebot einer Webseite aktueller beschreiben. Bei informationsorientierten Anfragen ist eine solche Aussage weniger eindeutig möglich. Einerseits, ist es für den Nutzer hilfreich zu sehen im welchem Umfeld die Suchbegriffe verwendet werden, andererseits ist aber auch der Gesamtkontext einer Seite nicht zu vernachlässigen, so dass sich hier die die Beschreibungsquellen gegenseitig ergänzen. Die Beschreibung aus den Metadaten bzw. aus einem Verzeichnis informiert den Nutzer über den Inhalt des Dokuments im Allgemeinen und der Auszug aus dem Inhalt, im welchem Kontext konkret das Suchargument vorkommt. Die Schwierigkeit in Zusammenhang mit den Anfragetypen besteht darin, dass die Möglichkeit für eine Suchmaschine den genauen Kontext einer Anfrage, und damit den Anfragetyp zu erfassen, sehr begrenzt sind. Eine denkbare Lösung könnte darin bestehen, sowohl eine allgemeine Beschreibung, als auch eine konkrete Angabe, in welchem Kontext der Suchbegriff im Inhalt vorzufinden ist, zu liefern. Unabhängig davon welche Beschreibungsquelle verwendet wird, ist die sprachliche Qualität und Vollständigkeit der Beschreibung nicht zu vernachlässigen, da sich die sprachliche Qualität einer Beschreibung entscheidend auf die Aussagekraft auswirken kann.

Unter Berücksichtigung der zuvor genannten Aspekte, sowie der Erkenntnisse der Studie, soll ein Schema zur Erstellung der Beschreibungen entwickelt werden, das im Folgenden erläutert werden soll.

Der Entwurf des Schemas geht von Grundannahme aus, dass einerseits von Menschen erstellte und andererseits sprachlich vollständige Beschreibungen zur Repräsentation von Texten in ihrem Gesamtkontext geeigneter sind, als Ausschnitte aus einem Text. So könnte im ersten Beschreibungselement eine Trefferbeschreibung aus den Metadaten oder den Verzeichnissen verwendet und diese durch einen Auszug aus dem Inhalt des Dokuments ergänzt werden. Dies ist zum Beispiel dann sinnvoll, wenn der Suchbegriff

nicht in der Beschreibung aus den erstgenannten Quellen vorkommt und damit der Zusammenhang mit der Suchanfrage möglicherweise nicht nachvollziehbar ist. Allerdings sollte die Anzahl der verwendeten Elemente möglichst gering gehalten werden, da sich eine Einbindung mehrerer Trefferbeschreibungen negativ auf die Aussagekraft auswirken kann, wie die Untersuchung gezeigt hat. Bei Google wurde in der Vergangenheit ein vergleichbares Verfahren eingesetzt, indem zusätzlich zum eigenen „Snippet“ eine Beschreibung aus dem ODP angezeigt wurde. Im Unterschied dazu erhalten die beschreibenden Quellen eine höhere Priorität als die KWIC Auszüge. Darüber hinaus wird die Beschreibung nur dann um ein weiteres Element ergänzt, das einen KWIC Auszug enthält, wenn anhand der Beschreibenden Quellen der Zusammenhang mit der Suchanfrage nicht nachvollziehbar ist.

Dies wäre mit einem Verfahren zu ergänzen, das einerseits die Länge der Beschreibung prüft und gegebenenfalls den Metadaten entnommenen Text bzw. den Ausschnitt aus dem Inhalt des Textes so kürzt, dass ein noch möglichst vollständiger Satz bzw. Satzausschnitt in der Beschreibung verbleibt. Abbildung 16 zeigt ein Ablaufdiagramm den vorgeschlagenen Schemas, das die Abfolge der einzelnen Schritte visualisiert.

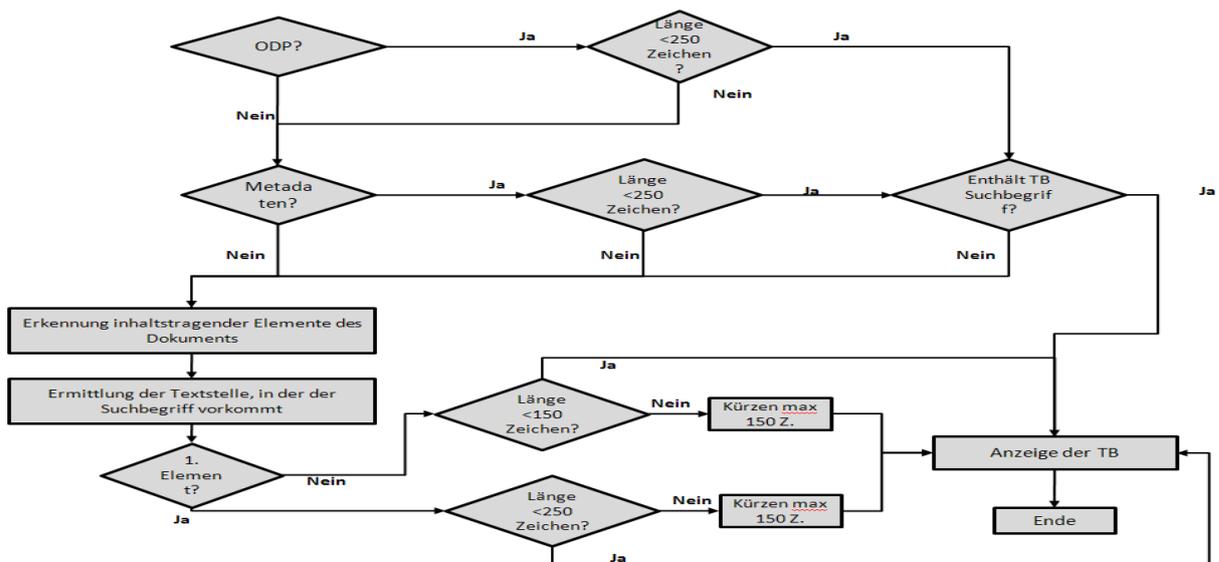


Abbildung 16 Ablaufdiagramm des Schemaentwurfs

Neben diesem grundsätzlichen Optimierungsfeld, das sich auf die Verwendung der verschiedenen Beschreibungsquellen bezieht, gibt es weitere Bereiche in denen zum Mindestens einige Suchmaschinen noch Nachholbedarf haben. Zum einem hat sich während der Analyse der Trefferbeschreibungen gezeigt, dass insbesondere Seekport noch massive Probleme hat, den eigentlichen Seiteninhalt zu erkennen. So wurden recht häufig ganze Menüelemente in die Trefferbeschreibungen übernommen, die als solche eine relativ geringe Aussagekraft haben. Verschiedene Möglichkeiten wie inhaltstragende Elemente einer Webseite zu identifizieren sind, werden von Lewandowski beschreiben (Lewandowski 2005, S. 220 ff.).

Ein anderer Bereich, in dem bei einigen Suchmaschinen ebenfalls Handlungsbedarf besteht, ist die effektive Erkennung von Dubletten bzw. von gleichlautenden Beschreibungen in den Metadaten. In solchen Fällen ist es möglicherweise sinnvoll, verstärkt auf Auszüge aus dem Text zurückzugreifen, damit die Unterschiede zwischen den einzelnen Seiten auf der Suchergebnisseite nachvollziehbar sind.

Auch bei Angeboten von Webseiten, deren Einzelseiten stets einer gleichen formalen Struktur folgen, wie zum Beispiel die Artikel von Wikipedia, ist noch Optimierungspotential zu sehen. So kann für solche Seiten ein Schema festgelegt werden, das besagt, dass für die Trefferbeschreibung immer ein bestimmter Bereich bzw. Abschnitt verwendet werden soll.

Aber auch Betreiber und Ersteller von Webseiten können Vorkehrungen treffen, um die Beschreibungen des eigenen Angebots zu verbessern. Dazu gehört die Bereitstellung von der Meta-Description, die insbesondere bei den Content Management Systemen auf die Inhalte der einzelnen Seiten angepasst werden sollten. Weiterhin ist eine kurze und prägnante Formulierung ratsam, damit die Beschreibung auch vollständig übernommen werden kann. Thurow empfiehlt in ihrem Buch „Search Engine Visibility“ eine Länge von 200 bis 250 Zeichen, was einer Länge von 2 bis 3 Zeilen entspricht (Thurow 2007, S. 73). Nicht direkt eine Optimierungsmöglichkeit, aber zum Mindestens eine Steuerungsmöglichkeit besteht darin, auf dem

Server eine robots.txt Datei zu hinterlegen, die die Angabe noodp enthält, die eine Verwendung von Beschreibungen aus dem ODP Verzeichnis verhindert (at-web 2008a). Allerdings ist fraglich, ob eine solche Unterbindung wünschenswert ist.

## 5 Fazit

Die zentralen Fragestellungen der Arbeit waren einerseits die derzeitige Verwendung der Trefferbeschreibungen sowie deren Struktur und andererseits die Beurteilung der Beschreibungstexte durch den Nutzer. In Bezug auf die Verwendung konnten teilweise sehr unterschiedliche Strategien seitens der Suchmaschinenbetreiber beobachtet werden. Jedoch steht außer Frage, dass derzeit die Auszüge aus dem Inhalt am häufigsten verwendet werden. Inwieweit dies auf das mangelnde Vorhandensein anderer Beschreibungsquellen, die jedoch bei der Bewertung besser als der KWIC-Ansatz beurteilt wurden, zurückzuführen ist, konnte im Rahmen dieser Studie nicht festgestellt werden.

Offensichtlich ist, dass einerseits eine sprachlich hohe Qualität und andererseits ein erkennbarer Zusammenhang zwischen der Suchanfrage und der Trefferbeschreibung gewünscht werden. Da die Anzahl der Treffer, die zu einer Fragestellung passen, in Zukunft und damit vergleichbare Webseiten eher zu nehmen werden, ist es umso wichtiger, dem Nutzer mit guten Beschreibungen, die den Inhalt eines Dokuments widerspiegeln, bei der Auswahl der in Frage kommenden Treffer zu unterstützen.

Inwieweit neuere Technologien, wie z.B. das „Semantic Web“, die Erstellung der Beschreibungen unterstützen kann, ist fraglich, da es sich dabei nicht um eine natürliche Sprache handelt, sondern der Ansatz verfolgt wird, über eine künstliche Sprache den semantischen Zusammenhang bzw. Inhalt einer Seite für eine Maschine verständlich zu machen. Allerdings könnten solche Technologien die Suchmaschine unterstützen, den Kontext einer Seite zu erfassen. Durch einen Abgleich zwischen den „maschinenverständlichen“ semantischen Beschreibungen und den „natürlichsprachigen“ Beschreibungen, könnte jene ausgewählt werden, die den Inhalt am passendsten repräsentiert.

Am vielversprechendsten scheint jedoch eine Integration des ODP-Ansatzes zu sein. So wäre zum Beispiel denkbar, dass zu jedem Treffer von dem Nut-

zer Vorschläge gemacht werden können, die anschließend für eine Beschreibung eines Treffers verwendet werden können.

Durch die Einbindung weiterer Kollektionen (die in der Regel präzisere Angaben enthalten) könnten auch reguläre Treffer Profit schlagen, wenn die Webtreffer durch die in Kollektionen befindlichen Angaben ergänzt werden<sup>17</sup>. Sinnvoll erscheint dies insbesondere bei solchen Kollektionen, in denen standardisierte Angaben zu einem Dokument erfasst werden, wie z.B. Büchersuche, oder die Videosuche. Allerdings sollte dies nur dann geschehen, wenn durch eine zusätzliche Beschreibung ein Mehrwert gewonnen wird, oder durch einen vollständigen Ersatz einer anderen Beschreibungsquelle der Informationsgehalt effektiv steigt.

Wie der letzte Abschnitt der Arbeit gezeigt hat, wird derzeit insgesamt noch ein deutliches Optimierungspotential bei allen Suchmaschinen gesehen, zum einem was die effektive Auswahl von geeigneten Quellen betrifft, und zum anderen, was die Vermeidung von Beschreibungs-Doubletten betrifft.

---

<sup>17</sup> Voraussetzung für eine solche Ergänzung ist jedoch, dass in einem vertikalen Suchraum eine zum Treffer gehörende Beschreibung vorliegt.

## 6 Literatur und Quellenverzeichnis

Alby u.a. 2007

ALBY, Tom ; KARZAUNINKAT, Stefan: *Suchmaschinenoptimierung : Professionelles Website-Marketing für besseres Ranking ; [berücksichtigt Google Analytics]*. 2., aktualisierte Aufl. München : Hanser, 2007.

Aschoff u.a. 2007

ASCHOFF, Franz. ; RAUSCH VON TRAUBENBERG, Eva: *Usability von Webportalen und Web- Suchmaschinen im Vergleich*. In: *Information Wissenschaft und Praxis* 58 (2007), Nr. 3, S. 141–148.

at-web 2004

AT-WEB: *Suchmaschine Seekport*

URL: <http://www.at-web.de/seekport/index.htm>.

Aktualisierungsdatum: 2004-11-08. – Überprüfungsdatum 2008-06-16

at-web 2008a

AT-WEB: *Suchmaschinen - Mehr Gemeinsamkeiten für Robots-Richtlinien*

URL: <http://www.at-web.de/blog/20080612/suchmaschinen-mehr-gemeinsamkeiten-fur-robots-richtlinien.htm>

Aktualisierungsdatum: 2008-06-12. – Überprüfungsdatum 2008-08-03

at-web 2008b

AT-WEB: *Interview mit Andreas Krawczyk, Head of Programming Yahoo! Deutschland*

URL: <http://www.at-web.de/blog/20080606/interview-mit-andreas-krawczyk-head-of-programming-yahoo-deutschland.htm>

Aktualisierungsdatum: 2008-06-06. – Überprüfungsdatum 2008-06-28

Beiler 2005

BEILER, Markus: Selektionsverhalten in den Ergebnislisten von Suchmaschinen. Modellentwicklung und empirische Überprüfung am Beispiel von Google. In: MACHILL, Marcel; SCHNEIDER, Norbert (Hrsg.). *Suchmaschinen: Neue Herausforderungen für die Medienpolitik*. Berlin : Vistas, 2005. (Schriftenreihe Medienforschung der Landesanstalt für Medien Nordrhein-Westfalen, 50), S. 165–189.

Crystal u.a. 2006

CRYSTAL, Abe ; GREENBERG, Jane: *Relevance Criteria Identified by Helth Information Users During Web Searches*. In: *Journal of the American Society for Information Science and Technology (JASIST)* 57 (2006), Nr. 10, S. 1368–1382.

Comscore 2008

COMSCORE: *comScore Veröffentlicht Europäische Rangliste zur Suche im Internet im März 2008: comScore Veröffentlicht Europäische Rangliste zur Suche im Internet im März 2008*. London, 2008-05-08

URL: <http://www.comscore.com/press/release.asp?press=2211>

Aktualisierungsdatum: 2008-05-08. – Überprüfungsdatum 2008-07-02

Drori 2000

DRORI, Offer: The Benefits of Displaying Additional Internal Document Information on Textual Database Search Result Lists. In: BORBINHA, José (Hrsg.). *Research and advanced technology for digital libraries : 4th European conference ; Lisbon, Portugal, September 18 - 20, 2000 ; proceedings*. Berlin : Springer, 2000, S. 69–82.

Dublin Core Metadata Initiative 1998

DUBLIN CORE METADATA INITIATIVE. *Dublin Core Metadata Element Set, Version 1.0: Reference Description*.

URL <http://dublincore.org/documents/1998/09/dces/#>

Aktualisierungsdatum: 1998-09. – Überprüfungsdatum 2008-08-11

Erlhofer 2007a

ERLHOFER, Sebastian: *Informationssuche im World Wide Web : Taktiken und Strategien bei der Nutzung von Suchmaschinen*. Berlin : wvb, 2007.

Erlhofer 2007b

ERLHOFER, Sebastian: *Suchmaschinen-Optimierung : Grundlagen, Funktionsweisen und Ranking-Optimierung ; [für Webentwickler ; Funktionsweisen von Google & Co. ; Ranking-Optimierung und usability ; inkl. TYPO3, WordPress und Web 2.0]*. 3., aktualisierte und erw. Aufl. Bonn : Galileo Press, 2007.

Fauldrath u.a. 2005

FAULDRATH, J. ; KUNISCH, A.: *Kooperative Evaluation der Usability von Suchmaschineninterfaces. Information*. In: *Information Wissenschaft und Praxis* 56 (2005), Nr. 1, S. 21–28.

Griesbaum 2004

GRIESBAUM, Joachim: *Evaluation of three German search engines : Altavista.de, Google.de and Lycos.de*. In: *IR informationResearch* 9 (2004), Nr. 4.

URL: <http://informationr.net/ir/9-4/paper189.html>

Haynes 2004

HAYNES, David: *Metadata for information management and retrieval*. London : Facet, 2004 .

Hearst 2006

HEARST, Marti A.: User Interfaces and Visualization. In: Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier (Hrsg.). *Modern information retrieval*. [Nachdr.]. Harlow : Pearson Addison-Wesley [u.a.], 2006, S. 257–323.

Hölscher 2002

HÖLSCHER, Christoph: *Die Rolle des Wissens im Internet : Gezielt suchen und kompetent auswählen*. Stuttgart : Klett-Cotta, 2002.

Kaczmarek 2006

KACZMIREK, Lars: *Information und Selektion : Gebrauchstauglichkeit der Ergebnisseiten von Suchmaschinen*. Mannheim : Universitätsbibliothek Mannheim (MADOC), 2006 (Mannheimer sozialwissenschaftliche Abschlussarbeiten 003/2006).

Lewandowski 2005

LEWANDOWSKI, Dirk: *Web Information Retrieval : Technologien zur Informationssuche im Internet*. Frankfurt am Main : Dt. Ges. für Informationswiss. und Informationspraxis, 2005.

Lewandowski 2007

LEWANDOWSKI, Dirk: Trefferpräsentation in Web-Suchmaschinen. In: OCKENFELD, M. (Hrsg.) (Hrsg.). *Information in Wissenschaft, Bildung und Wirtschaft : Proceedings der 29. Online-Tagung der DGI 2007*. Frankfurt am Main : DGI, 2007, S. 83–90.

Lewandowski 2008

LEWANDOWSKI, Dirk: *The Retrieval Effectiveness of Web Search Engines : Considering Results Descriptions*. In: *Journal of Documentation* 64 (2008), Preprint

URL: [http://www.durchdenken.de/lewandowski/doc/JDoc2008\\_preprint.pdf](http://www.durchdenken.de/lewandowski/doc/JDoc2008_preprint.pdf)

Machill u.a. 2003

MACHILL, Marcel u.a.: Wegweiser im Netz : Qualität und Nutzung von Suchmaschinen. In: MACHILL, Marcel (Hrsg.). *Wegweiser im Netz : Qualität und Nutzung von Suchmaschinen*. Gütersloh : Verl. Bertelsmann-Stiftung, 2003, S. 13–490.

Marchionini 1995

MARCHIONINI, Gary: *Information seeking in electronic environments*. Reprinted Cambridge : Cambridge Univ. Press, 1995.

Open Directory Project 2004

OPEN DIRECTORY PROJECT. *Editierstil-Richtlinien*

URL: <http://www.dmoz.org/World/Deutsch/guidelines/describing.html>

Aktualisierungsdatum: 2004-09-30. Überprüfungsdatum 2008-07-14

Open Directory Project 2008a

OPEN DIRECTORY PROJECT. *Über das Open Directory Project*. – Open Directory Project

URL: <http://www.dmoz.org/World/Deutsch/about.html>

Aktualisierungsdatum: 2006-09-18. – Überprüfungsdatum 2008-08-11

Open Directory Project 2008b

Open Directory Project. *Who We Are And What We Do*

URL: <http://www.dmoz.org/help/geninfo.html>. –

Aktualisierungsdatum: 2008-02-06. – Überprüfungsdatum 2008-08-11.

Rieh 2002

RIEH, Soo Young: *Judgement of Information Quality and Cognitive Authority in the Web*. In: *Journal of the American Society for Information Science and Technology (JASIST)* 53 (2002), Nr. 2, S. 145–161.

Schmidt-Mänz 2007

SCHMIDT-MÄNZ, Nadine: *Untersuchung des Suchverhaltens im Web : Interaktion von Internetnutzern mit Suchmaschinen*. Hamburg : Kovac, 2007.

Tang u.a. 2001

TANG, Rong ; SOLOMON, Paul: *Use of Relevance Criteria across Stages of Document Evaluation : On the Complementarity of Experimental and Naturalistic Studies*. In: *Journal of the American Society for Information Science and Technology (JASIST)* 52 (2001), Nr. 8, S. 676–685.

Thurrow 2007

THUROW, Shari: *Search engine visibility*. 2nd ed. Indianapolis: New Riders, 2007.

Yahoo Deutschland 2008

YAHOO DEUTSCHLAND. *Was ist das Yahoo! Verzeichnis?*

URL: <http://help.yahoo.com/l/de/yahoo/dir/dir001.html>

Aktualisierungsdatum: 2008-06-15. – Überprüfungsdatum 2008-06-15

van Eimeren u.a. 2008

VAN EIMEREN, Birgit ; FRESS, Beate: *Internetverarbeitung: Größter Zuwachs bei Silver-Surfern : Ergebnisse der ARD/ZDF Onlinestudie 2008*. In: *Media Perspektiven* 13 (2008), Nr. 7, S. 330–344.

Webhits 2008

WEBHITS. *Nutzung von Suchmaschinen*

URL: <http://www.webhits.de/deutsch/index.shtml?webstats.html>

Aktualisierungsdatum: 2008-06-16. – Überprüfungsdatum 2008-06-16

Wikimedia Foundation 2008

WIKIMEDIA FOUNDATION.: *Wie\_schreibe\_ich\_gute\_Artikel*

URL:[http://de.wikipedia.org/wiki/Wikipedia:Wie\\_schreibe\\_ich\\_gute\\_Artikel](http://de.wikipedia.org/wiki/Wikipedia:Wie_schreibe_ich_gute_Artikel)

Aktualisierungsdatum: 2008-07-05. – Überprüfungsdatum 2008-07-12

Wolf 2000

WOLF, C.: Vergleichende Evaluierung von Such- und Metasuchmaschinen.

*Proceedings of 7. Internationales Symposium für Informationswissenschaft,*

*Darmstadt, Germany.* Univ.-Verl. Konstanz, 2000, S. 31–48.

# Anhang

## A1. Suchanfragen

|    |                            |    |                    |
|----|----------------------------|----|--------------------|
| ID | Suchanfrage                | 44 | finanzen           |
| 1  | www.clipfish.de            | 45 | telefonauskunft    |
| 2  | fotoservice                | 46 | autokauf           |
| 3  | kijiji                     | 47 | xbox 360           |
| 4  | focus                      | 48 | spiegel            |
| 5  | opusforum                  | 49 | ing diba           |
| 6  | stadtplan                  | 50 | kaufvertrag        |
| 7  | bigbrother<br>www.deutsche | 51 | google.de          |
| 8  | bank.d                     | 52 | meinstadt.de       |
| 9  | www.dasoertliche.de        | 53 | dove               |
| 10 | weltbild.de                | 54 | cheats             |
| 11 | studivz                    | 55 | adaware            |
| 13 | google.com                 | 56 | parship.de         |
| 14 | amazonen                   | 58 | uno                |
| 15 | eu-neuwagen                | 59 | www.radio7.de      |
| 16 | großarl                    | 60 | www.tomtom.com     |
| 18 | tchibo reisen              | 61 | toom               |
| 19 | aktion mensch              | 62 | goldesel           |
| 20 | babywalz                   | 63 | dab bank           |
| 21 | automarken                 | 64 | antolin.de         |
| 22 | www.schlach.com            | 65 | sparda hessen      |
| 23 | laferlichterlecker         | 66 | freeware           |
| 24 | s.oliver                   | 67 | discount 24        |
| 25 | air berlin                 | 68 | www.sonnenklar.de  |
| 26 | www.lufthansa.com          | 69 | globetrotter       |
| 27 | okf                        | 70 | t-online           |
| 29 | nokia 6233                 | 72 | spinchat.de        |
| 30 | saugstube                  | 73 | www.aldi.com       |
| 31 | weight watchers            | 74 | unfallversicherung |
| 32 | knut                       | 75 | musik              |
| 33 | jahreswagen                | 76 | premiere           |
| 34 | mcdonalds                  | 77 | bahn               |
| 35 | my-video                   | 78 | umzüge             |
| 36 | telefonrechnung            | 79 | www.               |
| 37 | leipzig                    | 80 | www.arbeitsagentur |
| 38 | werder bremen              | 81 | www.toggo.de       |
| 39 | conrad                     | 82 | sparkasse neuss    |
| 40 | onvista                    | 83 | germanwings        |
| 41 | o2                         | 84 | last minute        |
| 42 | saturn.de                  | 86 | flatcast           |
| 43 | paris hilton               | 87 | lafée              |
|    |                            | 88 | aachener zeitung   |

89 wetter  
90 freemail  
91 segmüller  
92 gmx  
93 dubai  
94 aooa  
95 dresdner bank  
96 lycos  
97 autoscout  
98 mein video  
102 netdokter  
103 domain  
104 www.e-bay.de  
105 immobilien scout  
106 wetter.com  
107 media markt  
108 goo  
111 rasenmäher  
112 saturn  
113 böhse onkelz

## A2. Fragebogen<sup>18</sup>

Hier die wichtigsten Informationen zum Ausfüllen des Fragebogens:

- **Die Beschreibungen sollen in Bezug auf ihre Aussagekraft bewertet werden.**  
**Das heißt, es geht darum, wie aussagekräftig die Trefferbeschreibung den zu erwartenden Inhalt einer Webseite beschreibt.**
- **Die Treffer sollen einzeln bewertet werden**

Es geht **NICHT** darum:

- a) Eine Relevanzbeurteilung vorzunehmen (d.h. zu bewerten, ob ein Treffer zu einem Thema passt).
- b) Die Treffer untereinander zu vergleichen.

Der Fokus liegt auf dem Beschreibungstext unterhalb des Titels(im Beispiel umrandeter Bereich)

### Joomla-Workshop | Startseite

... Joomla! ... Im Joomla! ... wohl kaum einem Joomla! ... Da aber sicherlich nicht alle Joomla! ...  
Sicherlich ist es niemand aus der Joomla! ... Joomla! ... Antwort: Im Netz gibt es jede Menge Joomla!  
... " wird Euch bei der Arbeit mit Joomla! ... Checkliste: Neue Joomla! ... Joomla! ... Joomla! ... Namen  
der Joomla! ... PHP-Befehle im Joomla! ... Joomla! ... Joomla! ...

Die Aussagekraft der Beschreibungen soll auf einer Skala von 1 bis 6 bewertet werden, wobei diese Skala den deutschen Schulnoten entspricht.

**1 = sehr gut**

**2 = gut**

**3 = zufriedenstellend**

**4 = ausreichend**

**5 = mangelhaft**

**6 = ungenügend**

---

<sup>18</sup> Der Bewertungsteil des Fragebogens enthält nur die Seiten 1 und 3 eines Fragebogensatzes statt der 10 Seiten, die ein Satz enthält.

### **Witzige Videoclips - Dein Witzvideo - Gratis Witzig-Videos bei Clipfish**

Witzige Videoclips kostenlos! Sehen, lachen, mitmachen – lade Witzig-Videos ... © Clipfish.de - Witzige Videoclips - Dein Witze-Video und mehr .

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

3382

### **azrael74.de » Clipfish**

azrael74.de / Clipfish / 7. August, 2006 @ 12:11 · Abgelegt unter Netzwelt / Auch RTL springt jetzt auf den Video-Zug auf. "Clipfish" heißt das neueste Baby aus dem Hause RTL Interactive. Permalink / www.flickr.com / 1 Kommentar » / clipfish - metaspinner.de Said, / 3. September, 2007 @ 15:01 / [...] [...] ...

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

2294

### **Bianca Ryan - A star is born - talents bianca ryan star - Video**

Funny Clips kostenlos! Sehen, lachen, mitmachen: lade dein funny Video hoch und ... © Clipfish.de - Bianca Ryan - A star is born-Video - Video von hannelena - Lustige ...

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

7034

### **videograb.de - Videos von YouTube, Google.Video, MyVideo ...**

z.B. <http://www.clipfish.de/player.php?videoid> ...4 Min. und 8 Sek.

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

1270

### **onlinejournalismus.de - Das Maqazin zum Thema**

BR-Online in frischem Gewand: Mia gfoits ... Dabei soll ebenfalls ein Zusammenhang zu dem Werbeumfeld, also den Online Videos auf Webseiten wie beispielsweise RTLnow, YouTube, Clipfish, Spiegel.de, Zeit.de, hergestellt werden. " Hier geht es zu der Umfrage: [www.unipark.de/uc/k\\_rfh\\_medien/50da](http://www.unipark.de/uc/k_rfh_medien/50da). ...

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

1301

### **ALDI Fotoservice: Online Bilderdienst, Online Fotoalbum, Online ...**

ALDI Foto Service: Foto Manager, Online Foto Album und Online Druckservice - Die Rundumlösung für Ihre digitalen Fotos. Nutzen Sie unseren Fotoservice im ...

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

8646

### **Fotoservice-Vergleich**

Digitalfotos entwickeln - Fotoservice-Preisrechner, Anleitungen und Informationen zur Digitalfotografie...

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

7799

### **T-Online Fotoservice**

Fotobuch. Testergebnis "gut" in Ausgabe 6/2007 der Computer Bild. Ihre Fotos im Digitaldruck. Umschlag: Hard- oder Softcover. 3 Formate, bis zu 98 Seiten

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

2717

### **Fotoservice**

Fotoservice. Bestellen ist ... Fotoservice Software. Fotoabzüge. Fotoabzüge. Fotoabzüge ... mehr zu Fotoservice Software. Shopping-Tipps: Digitalkameras ...

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

257

### **MMS - Fotoservice | T-Mobile**

T-Mobile MMS - Fotoservice So funktioniert es - Mit dem Fotoservice von T-Mobile und KODAK Bilder speichern, drucken und online zeigen

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

7508

Welche Suchmaschine wird von dir bevorzugt?

- Ask
- Google
- MSN/Live
- Seekport
- Yahoo
- \_\_\_\_\_

Was ist dir bei einer Trefferbeschreibung besonders wichtig? (eine Antwort)

- Beschreibung in ganzen Sätzen
- Ausschnitt aus dem Inhalt der Seite
- Ausschnitt der Textstellen, in denen die Suchwörter vorkommen
- Anzeige von einzelnen Stichwörtern / Kategorien

Wie viele Zeilen sollte eine ideale Beschreibung haben?

- 1 Zeile
- 2 Zeilen
- 3 Zeilen
- 4 Zeilen
- 5 Zeilen und mehr

Vielen Dank für das Ausfüllen des Fragebogens!

Lukas Tittmann

### A3. Zusammensetzung der Bewertungssets

Die Sets zu jeder Suchanfrage wurden in fünf Teile aufgeteilt und den Gruppen A bis E zugeordnet. A entspricht den ersten zwei Seiten und E den letzten zwei Seiten, also den Seiten 9 und 10. Anschließend wurden in einer Tabelle die Gruppen A und C aufsteigend und B und D absteigend angeordnet. In einem weiteren Schritt wurden in den Gruppen C und D um jeweils die Hälfte getauscht, sprich Position 1-51 befindet sich nun an Position 52-102 und Position 52-102 entsprechend bei 1-51. Die Gruppe E wurde zufällig getauscht. In jenen Sets, in denen eine Suchanfrage zwei mal vorkommt, wurde eine der Suchanfragen mit dem Nachbarset getauscht.

Folgende Tabelle gibt detaillierte Auskunft über die Anordnung der Sets: SA gibt die Nummer der Suchanfrage an (s. Anhang A1). Die Seitenspalte gibt die Seitenzahl der ersten Seite des Seitenpaares an.

|    | A (1+2) |       | B (3+4) |       | C (5+6) |       | D (7+8) |       | E (9+10) |       |
|----|---------|-------|---------|-------|---------|-------|---------|-------|----------|-------|
|    | SA      | Seite | SA      | Seite | SA      | Seite | SA      | Seite | SA       | Seite |
| 1  | 1       | 2     | 113     | 1002  | 55      | 514   | 54      | 506   | 2        | 20    |
| 2  | 2       | 12    | 112     | 992   | 56      | 524   | 53      | 496   | 6        | 60    |
| 3  | 3       | 22    | 111     | 982   | 58      | 534   | 52      | 486   | 87       | 801   |
| 4  | 4       | 32    | 108     | 972   | 59      | 544   | 51      | 476   | 78       | 723   |
| 5  | 5       | 42    | 107     | 962   | 60      | 554   | 50      | 466   | 45       | 418   |
| 6  | 6       | 52    | 106     | 952   | 61      | 564   | 49      | 456   | 55       | 518   |
| 7  | 7       | 62    | 105     | 942   | 62      | 574   | 48      | 446   | 67       | 628   |
| 8  | 8       | 72    | 104     | 935   | 63      | 584   | 47      | 436   | 94       | 871   |
| 9  | 9       | 81    | 103     | 925   | 64      | 594   | 46      | 426   | 42       | 388   |
| 10 | 10      | 91    | 102     | 915   | 65      | 604   | 45      | 416   | 80       | 741   |
| 11 | 11      | 101   | 98      | 904   | 66      | 614   | 44      | 406   | 37       | 338   |
| 12 | 13      | 111   | 97      | 895   | 67      | 624   | 43      | 396   | 47       | 438   |
| 13 | 14      | 121   | 96      | 885   | 68      | 634   | 42      | 386   | 32       | 288   |
| 14 | 15      | 131   | 95      | 875   | 69      | 641   | 41      | 376   | 86       | 791   |
| 15 | 16      | 141   | 94      | 865   | 70      | 651   | 40      | 366   | 82       | 761   |
| 16 | 18      | 151   | 93      | 855   | 72      | 661   | 39      | 356   | 81       | 751   |
| 17 | 19      | 161   | 92      | 845   | 73      | 671   | 38      | 346   | 48       | 448   |
| 18 | 20      | 171   | 91      | 835   | 74      | 679   | 37      | 336   | 107      | 968   |
| 19 | 21      | 181   | 90      | 825   | 75      | 689   | 36      | 326   | 61       | 568   |
| 20 | 22      | 191   | 89      | 815   | 76      | 699   | 35      | 316   | 34       | 308   |
| 21 | 23      | 200   | 88      | 805   | 77      | 709   | 34      | 306   | 105      | 948   |
| 22 | 24      | 210   | 87      | 795   | 78      | 719   | 33      | 296   | 15       | 139   |
| 23 | 25      | 220   | 86      | 785   | 79      | 729   | 32      | 286   | 24       | 218   |

|    |    |            |    |            |     |             |     |             |     |            |
|----|----|------------|----|------------|-----|-------------|-----|-------------|-----|------------|
| 24 | 26 | <b>230</b> | 84 | <b>775</b> | 80  | <b>737</b>  | 31  | <b>276</b>  | 66  | <b>618</b> |
| 25 | 27 | <b>240</b> | 83 | <b>765</b> | 81  | <b>747</b>  | 30  | <b>266</b>  | 84  | <b>781</b> |
| 26 | 29 | <b>250</b> | 82 | <b>755</b> | 83  | <b>767</b>  | 29  | <b>256</b>  | 19  | <b>169</b> |
| 27 | 30 | <b>260</b> | 81 | <b>745</b> | 82  | <b>757</b>  | 27  | <b>246</b>  | 60  | <b>558</b> |
| 28 | 31 | <b>270</b> | 80 | <b>735</b> | 84  | <b>777</b>  | 26  | <b>236</b>  | 52  | <b>488</b> |
| 29 | 32 | <b>280</b> | 79 | <b>727</b> | 86  | <b>787</b>  | 25  | <b>226</b>  | 97  | <b>901</b> |
| 30 | 33 | <b>290</b> | 78 | <b>717</b> | 87  | <b>797</b>  | 24  | <b>216</b>  | 10  | <b>99</b>  |
| 31 | 34 | <b>300</b> | 77 | <b>707</b> | 88  | <b>807</b>  | 23  | <b>206</b>  | 8   | <b>80</b>  |
| 32 | 35 | <b>310</b> | 76 | <b>697</b> | 89  | <b>817</b>  | 22  | <b>197</b>  | 49  | <b>458</b> |
| 33 | 36 | <b>320</b> | 75 | <b>687</b> | 90  | <b>827</b>  | 21  | <b>187</b>  | 1   | <b>119</b> |
| 34 | 37 | <b>330</b> | 74 | <b>677</b> | 91  | <b>837</b>  | 20  | <b>177</b>  | 96  | <b>891</b> |
| 35 | 38 | <b>340</b> | 73 | <b>669</b> | 92  | <b>847</b>  | 19  | <b>167</b>  | 63  | <b>588</b> |
| 36 | 39 | <b>350</b> | 72 | <b>659</b> | 93  | <b>857</b>  | 18  | <b>157</b>  | 14  | <b>129</b> |
| 37 | 40 | <b>360</b> | 70 | <b>649</b> | 94  | <b>867</b>  | 16  | <b>147</b>  | 70  | <b>655</b> |
| 38 | 41 | <b>370</b> | 69 | <b>639</b> | 95  | <b>877</b>  | 15  | <b>137</b>  | 11  | <b>109</b> |
| 39 | 42 | <b>380</b> | 68 | <b>632</b> | 96  | <b>887</b>  | 14  | <b>127</b>  | 77  | <b>713</b> |
| 40 | 43 | <b>390</b> | 67 | <b>622</b> | 97  | <b>897</b>  | 13  | <b>117</b>  | 103 | <b>931</b> |
| 41 | 44 | <b>400</b> | 66 | <b>612</b> | 98  | <b>906</b>  | 11  | <b>107</b>  | 65  | <b>608</b> |
| 42 | 45 | <b>410</b> | 65 | <b>602</b> | 102 | <b>917</b>  | 10  | <b>97</b>   | 108 | <b>978</b> |
| 43 | 46 | <b>420</b> | 64 | <b>592</b> | 103 | <b>927</b>  | 9   | <b>87</b>   | 102 | <b>921</b> |
| 44 | 47 | <b>430</b> | 63 | <b>582</b> | 104 | <b>937</b>  | 8   | <b>78</b>   | 27  | <b>248</b> |
| 45 | 48 | <b>440</b> | 62 | <b>572</b> | 105 | <b>944</b>  | 7   | <b>68</b>   | 46  | <b>428</b> |
| 46 | 49 | <b>450</b> | 61 | <b>562</b> | 106 | <b>954</b>  | 6   | <b>58</b>   | 53  | <b>498</b> |
| 47 | 50 | <b>460</b> | 60 | <b>552</b> | 107 | <b>964</b>  | 5   | <b>48</b>   | 75  | <b>693</b> |
| 48 | 51 | <b>470</b> | 59 | <b>542</b> | 108 | <b>974</b>  | 4   | <b>38</b>   | 41  | <b>378</b> |
| 49 | 52 | <b>480</b> | 58 | <b>532</b> | 111 | <b>984</b>  | 3   | <b>28</b>   | 58  | <b>538</b> |
| 50 | 53 | <b>490</b> | 56 | <b>522</b> | 112 | <b>994</b>  | 2   | <b>18</b>   | 54  | <b>508</b> |
| 51 | 54 | <b>500</b> | 55 | <b>512</b> | 113 | <b>1004</b> | 1   | <b>8</b>    | 56  | <b>528</b> |
| 52 | 55 | <b>510</b> | 54 | <b>502</b> | 1   | <b>6</b>    | 113 | <b>1006</b> | 30  | <b>268</b> |
| 53 | 56 | <b>520</b> | 53 | <b>492</b> | 2   | <b>16</b>   | 110 | <b>996</b>  | 31  | <b>278</b> |
| 54 | 58 | <b>530</b> | 52 | <b>482</b> | 3   | <b>26</b>   | 111 | <b>986</b>  | 43  | <b>398</b> |
| 55 | 59 | <b>540</b> | 51 | <b>472</b> | 4   | <b>36</b>   | 108 | <b>976</b>  | 62  | <b>578</b> |
| 56 | 60 | <b>550</b> | 50 | <b>462</b> | 5   | <b>46</b>   | 107 | <b>966</b>  | 64  | <b>598</b> |
| 57 | 61 | <b>560</b> | 49 | <b>452</b> | 6   | <b>56</b>   | 106 | <b>956</b>  | 90  | <b>831</b> |
| 58 | 62 | <b>570</b> | 48 | <b>442</b> | 7   | <b>66</b>   | 105 | <b>946</b>  | 93  | <b>861</b> |
| 59 | 63 | <b>580</b> | 47 | <b>432</b> | 8   | <b>76</b>   | 104 | <b>939</b>  | 5   | <b>50</b>  |
| 60 | 64 | <b>590</b> | 46 | <b>422</b> | 9   | <b>85</b>   | 103 | <b>929</b>  | 72  | <b>665</b> |
| 61 | 65 | <b>600</b> | 45 | <b>412</b> | 10  | <b>95</b>   | 102 | <b>919</b>  | 110 | <b>998</b> |
| 62 | 66 | <b>610</b> | 44 | <b>402</b> | 11  | <b>105</b>  | 98  | <b>908</b>  | 16  | <b>149</b> |
| 63 | 67 | <b>620</b> | 43 | <b>392</b> | 13  | <b>115</b>  | 97  | <b>899</b>  | 20  | <b>179</b> |
| 64 | 68 | <b>630</b> | 42 | <b>382</b> | 14  | <b>125</b>  | 96  | <b>889</b>  | 21  | <b>189</b> |
| 65 | 69 | <b>637</b> | 41 | <b>372</b> | 15  | <b>135</b>  | 95  | <b>879</b>  | 35  | <b>318</b> |
| 66 | 70 | <b>647</b> | 40 | <b>362</b> | 16  | <b>145</b>  | 94  | <b>869</b>  | 44  | <b>408</b> |
| 67 | 72 | <b>657</b> | 39 | <b>352</b> | 18  | <b>155</b>  | 93  | <b>859</b>  | 76  | <b>703</b> |
| 68 | 73 | <b>667</b> | 38 | <b>342</b> | 19  | <b>165</b>  | 92  | <b>849</b>  | 51  | <b>478</b> |

|     |     |             |    |            |    |            |    |            |     |             |
|-----|-----|-------------|----|------------|----|------------|----|------------|-----|-------------|
| 69  | 74  | <b>675</b>  | 37 | <b>332</b> | 20 | <b>175</b> | 91 | <b>839</b> | 111 | <b>988</b>  |
| 70  | 75  | <b>685</b>  | 36 | <b>322</b> | 21 | <b>185</b> | 90 | <b>829</b> | 4   | <b>40</b>   |
| 71  | 76  | <b>695</b>  | 35 | <b>312</b> | 22 | <b>195</b> | 89 | <b>819</b> | 95  | <b>881</b>  |
| 72  | 77  | <b>705</b>  | 34 | <b>302</b> | 23 | <b>204</b> | 88 | <b>809</b> | 29  | <b>258</b>  |
| 73  | 78  | <b>715</b>  | 33 | <b>292</b> | 24 | <b>214</b> | 87 | <b>799</b> | 69  | <b>645</b>  |
| 74  | 79  | <b>725</b>  | 32 | <b>282</b> | 25 | <b>224</b> | 86 | <b>789</b> | 3   | <b>30</b>   |
| 75  | 80  | <b>733</b>  | 31 | <b>272</b> | 26 | <b>234</b> | 84 | <b>779</b> | 38  | <b>348</b>  |
| 76  | 81  | <b>743</b>  | 30 | <b>262</b> | 27 | <b>244</b> | 83 | <b>769</b> | 36  | <b>328</b>  |
| 77  | 82  | <b>753</b>  | 29 | <b>252</b> | 30 | <b>264</b> | 81 | <b>749</b> | 98  | <b>910</b>  |
| 78  | 83  | <b>763</b>  | 27 | <b>242</b> | 29 | <b>254</b> | 82 | <b>759</b> | 88  | <b>811</b>  |
| 79  | 84  | <b>773</b>  | 26 | <b>232</b> | 31 | <b>274</b> | 80 | <b>739</b> | 18  | <b>159</b>  |
| 80  | 86  | <b>783</b>  | 25 | <b>222</b> | 32 | <b>284</b> | 79 | <b>731</b> | 7   | <b>70</b>   |
| 81  | 87  | <b>793</b>  | 24 | <b>212</b> | 33 | <b>294</b> | 78 | <b>721</b> | 113 | <b>1008</b> |
| 82  | 88  | <b>803</b>  | 23 | <b>202</b> | 34 | <b>304</b> | 77 | <b>711</b> | 39  | <b>358</b>  |
| 83  | 89  | <b>813</b>  | 22 | <b>193</b> | 35 | <b>314</b> | 76 | <b>701</b> | 83  | <b>771</b>  |
| 84  | 90  | <b>823</b>  | 21 | <b>183</b> | 36 | <b>324</b> | 75 | <b>691</b> | 26  | <b>238</b>  |
| 85  | 91  | <b>833</b>  | 20 | <b>173</b> | 37 | <b>334</b> | 74 | <b>681</b> | 40  | <b>368</b>  |
| 86  | 92  | <b>843</b>  | 19 | <b>163</b> | 38 | <b>344</b> | 73 | <b>673</b> | 91  | <b>841</b>  |
| 87  | 93  | <b>853</b>  | 18 | <b>153</b> | 39 | <b>354</b> | 72 | <b>663</b> | 25  | <b>228</b>  |
| 88  | 94  | <b>863</b>  | 16 | <b>143</b> | 40 | <b>364</b> | 70 | <b>653</b> | 89  | <b>821</b>  |
| 89  | 95  | <b>873</b>  | 15 | <b>133</b> | 41 | <b>374</b> | 69 | <b>643</b> | 106 | <b>958</b>  |
| 90  | 96  | <b>883</b>  | 14 | <b>123</b> | 42 | <b>384</b> | 68 | <b>636</b> | 74  | <b>683</b>  |
| 91  | 97  | <b>893</b>  | 13 | <b>113</b> | 43 | <b>394</b> | 67 | <b>626</b> | 23  | <b>208</b>  |
| 92  | 98  | <b>902</b>  | 11 | <b>103</b> | 44 | <b>404</b> | 66 | <b>616</b> | 59  | <b>548</b>  |
| 93  | 102 | <b>913</b>  | 10 | <b>93</b>  | 45 | <b>414</b> | 65 | <b>606</b> | 1   | <b>10</b>   |
| 94  | 103 | <b>923</b>  | 9  | <b>83</b>  | 46 | <b>424</b> | 64 | <b>596</b> | 33  | <b>298</b>  |
| 95  | 104 | <b>933</b>  | 8  | <b>74</b>  | 47 | <b>434</b> | 63 | <b>586</b> | 9   | <b>89</b>   |
| 96  | 105 | <b>940</b>  | 7  | <b>64</b>  | 48 | <b>444</b> | 62 | <b>576</b> | 92  | <b>851</b>  |
| 97  | 106 | <b>950</b>  | 6  | <b>54</b>  | 49 | <b>454</b> | 61 | <b>566</b> | 50  | <b>468</b>  |
| 98  | 107 | <b>960</b>  | 5  | <b>44</b>  | 50 | <b>464</b> | 60 | <b>556</b> | 22  | <b>199</b>  |
| 99  | 108 | <b>970</b>  | 4  | <b>34</b>  | 51 | <b>474</b> | 59 | <b>546</b> |     |             |
| 100 | 111 | <b>980</b>  | 3  | <b>24</b>  | 52 | <b>484</b> | 58 | <b>536</b> |     |             |
| 101 | 112 | <b>990</b>  | 2  | <b>14</b>  | 53 | <b>494</b> | 56 | <b>526</b> |     |             |
| 102 | 113 | <b>1000</b> | 1  | <b>4</b>   | 54 | <b>504</b> | 55 | <b>516</b> |     |             |

#### A4. Paarungen von Beschreibungstypen im 1. Und

##### 2. Element

| Typ            | 2 Be-<br>schrtyp | Ask            | Google         | MSN/Live       | Seekport       | Yahoo          | Alle SM        |
|----------------|------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| KWIC           | KWIC             | 70,86%         | 75,00%         | 8,96%          | 92,20%         | 44,98%         | <b>63,88%</b>  |
|                | Metadaten        | 0,57%          | 3,38%          | 0,00%          | 0,13%          | 0,28%          | <b>0,44%</b>   |
|                | ODP              | 0,00%          | 1,35%          | 0,00%          | 0,00%          | 0,00%          | <b>0,10%</b>   |
|                | Sonstiges        | 1,14%          | 0,00%          | 0,94%          | 0,38%          | 0,57%          | <b>0,54%</b>   |
|                | <b>Summe</b>     | <b>72,57%</b>  | <b>79,73%</b>  | <b>9,91%</b>   | <b>92,71%</b>  | <b>45,83%</b>  | <b>64,97%</b>  |
| Meta-<br>daten | KWIC             | 18,86%         | 16,22%         | 83,02%         | 0,26%          | 48,23%         | <b>28,46%</b>  |
|                | Metadaten        | 1,71%          | 0,00%          | 2,36%          | 0,00%          | 0,99%          | <b>0,74%</b>   |
|                | Sonstiges        | 0,00%          | 0,68%          | 0,94%          | 0,00%          | 0,99%          | <b>0,49%</b>   |
|                | <b>Summe</b>     | <b>20,57%</b>  | <b>16,89%</b>  | <b>86,32%</b>  | <b>0,26%</b>   | <b>50,21%</b>  | <b>29,69%</b>  |
| ODP            | KWIC             | 1,71%          | 0,00%          | 0,00%          | 0,00%          | 0,28%          | <b>0,25%</b>   |
| Sonstiges      | KWIC             | 2,86%          | 3,38%          | 2,83%          | 6,27%          | 2,69%          | <b>4,15%</b>   |
|                | Sonstiges        | 2,29%          | 0,00%          | 0,94%          | 0,77%          | 0,99%          | <b>0,94%</b>   |
|                | <b>Summe</b>     | <b>5,14%</b>   | <b>3,38%</b>   | <b>3,77%</b>   | <b>7,03%</b>   | <b>3,68%</b>   | <b>5,09%</b>   |
|                |                  | <b>100,00%</b> | <b>100,00%</b> | <b>100,00%</b> | <b>100,00%</b> | <b>100,00%</b> | <b>100,00%</b> |
| <b>N</b>       |                  | <b>175</b>     | <b>148</b>     | <b>212</b>     | <b>782</b>     | <b>707</b>     | <b>2024</b>    |

Anteile der Treffer, in denen im 1. oder 2. Element KWIC vorkommt:

Berücksichtigt werden nur jene Treffer, die auch einen Beschreibungstext enthalten.

|                     | => 2 Elemente | Alle Elemente |
|---------------------|---------------|---------------|
| KWIC im 1. Element: | 1315          | 2556          |
| KWIC im 2. Element: | 665           | 665           |
| <b>Summe</b>        | <b>1980</b>   | <b>3229</b>   |
| N                   | 2024          | 4912          |
| Anteil KWIC an N    | <b>97,83%</b> | <b>65,61%</b> |

## A5. Länge der Trefferbeschreibungen (in Wörter)

Anzahl

|        |         | Suchmaschine |        |          |          |       | Alle SM |
|--------|---------|--------------|--------|----------|----------|-------|---------|
|        |         | Ask          | Google | MSN/Live | Seekport | Yahoo |         |
|        | <= 5    | 43           | 18     | 26       | 10       | 15    | 112     |
|        | 6 - 10  | 105          | 96     | 72       | 14       | 25    | 312     |
|        | 11 - 15 | 239          | 164    | 152      | 11       | 51    | 617     |
|        | 16 - 20 | 455          | 387    | 288      | 13       | 346   | 1.489   |
|        | 21 - 25 | 166          | 304    | 306      | 24       | 426   | 1.226   |
|        | 26 - 30 | 2            | 37     | 116      | 59       | 75    | 289     |
|        | 31 - 35 | 0            | 0      | 10       | 117      | 0     | 127     |
|        | 36 - 40 | 0            | 0      | 0        | 222      | 0     | 222     |
|        | 41 - 45 | 0            | 0      | 1        | 229      | 0     | 230     |
|        | 46 - 50 | 0            | 0      | 0        | 180      | 0     | 180     |
|        | 51+     | 0            | 0      | 0        | 108      | 0     | 108     |
| Gesamt |         | 1.010        | 1.006  | 971      | 987      | 938   | 4.912   |

Anteile

|        |         | Suchmaschine |         |          |          |         | Alle SM |
|--------|---------|--------------|---------|----------|----------|---------|---------|
|        |         | Ask          | Google  | MSN/Live | Seekport | Yahoo   |         |
|        | <= 5    | 4,26%        | 1,79%   | 2,68%    | 1,01%    | 1,60%   | 2,28%   |
|        | 6-10    | 10,40%       | 9,54%   | 7,42%    | 1,42%    | 2,67%   | 6,35%   |
|        | 11-15   | 23,66%       | 16,30%  | 15,65%   | 1,11%    | 5,44%   | 12,56%  |
|        | 16 - 20 | 45,05%       | 38,47%  | 29,66%   | 1,32%    | 36,89%  | 30,31%  |
|        | 21 - 25 | 16,44%       | 30,22%  | 31,51%   | 2,43%    | 45,42%  | 24,96%  |
|        | 26 - 30 | 0,20%        | 3,68%   | 11,95%   | 5,98%    | 8,00%   | 5,88%   |
|        | 31 - 35 | 0,00%        | 0,00%   | 1,03%    | 11,85%   | 0,00%   | 2,59%   |
|        | 36 - 40 | 0,00%        | 0,00%   | 0,00%    | 22,49%   | 0,00%   | 4,52%   |
|        | 41 - 45 | 0,00%        | 0,00%   | 0,10%    | 23,20%   | 0,00%   | 4,68%   |
|        | 46 - 50 | 0,00%        | 0,00%   | 0,00%    | 18,24%   | 0,00%   | 3,66%   |
|        | 51+     | 0,00%        | 0,00%   | 0,00%    | 10,94%   | 0,00%   | 2,20%   |
| Gesamt |         | 100,00%      | 100,00% | 100,00%  | 100,00%  | 100,00% | 100,00% |

## A6.Länge der Trefferbeschreibungen (in Zeichen)

Anzahl

|        |           | Suchmaschine |        |          |          |       | Alle SM |
|--------|-----------|--------------|--------|----------|----------|-------|---------|
|        |           | Ask          | Google | MSN/Live | Seekport | Yahoo |         |
|        | <= 40     | 35           | 9      | 25       | 9        | 11    | 89      |
|        | 41 - 80   | 86           | 84     | 53       | 13       | 27    | 263     |
|        | 81 - 120  | 234          | 112    | 156      | 8        | 23    | 533     |
|        | 121 - 160 | 653          | 792    | 301      | 9        | 550   | 2.305   |
|        | 161 - 200 | 2            | 9      | 434      | 14       | 326   | 785     |
|        | 201 - 240 | 0            | 0      | 2        | 29       | 1     | 32      |
|        | 241 - 280 | 0            | 0      | 0        | 80       | 0     | 80      |
|        | 281 - 320 | 0            | 0      | 0        | 390      | 0     | 390     |
|        | 321 - 360 | 0            | 0      | 0        | 400      | 0     | 400     |
|        | 361+      | 0            | 0      | 0        | 35       | 0     | 35      |
| Gesamt |           | 1.010        | 1.006  | 971      | 987      | 938   | 4.912   |

Anteile

|        |           | Suchmaschine |         |          |          |         | Alle SM |
|--------|-----------|--------------|---------|----------|----------|---------|---------|
|        |           | Ask          | Google  | MSN/Live | Seekport | Yahoo   |         |
|        | <= 40     | 3,47%        | 0,89%   | 2,57%    | 0,91%    | 1,17%   | 1,81%   |
|        | 41 - 80   | 8,51%        | 8,35%   | 5,46%    | 1,32%    | 2,88%   | 5,35%   |
|        | 81 - 120  | 23,17%       | 11,13%  | 16,07%   | 0,81%    | 2,45%   | 10,85%  |
|        | 121 - 160 | 64,65%       | 78,73%  | 31,00%   | 0,91%    | 58,64%  | 46,93%  |
|        | 161 - 200 | 0,20%        | 0,89%   | 44,70%   | 1,42%    | 34,75%  | 15,98%  |
|        | 201 - 240 | 0,00%        | 0,00%   | 0,21%    | 2,94%    | 0,11%   | 0,65%   |
|        | 241 - 280 | 0,00%        | 0,00%   | 0,00%    | 8,11%    | 0,00%   | 1,63%   |
|        | 281 - 320 | 0,00%        | 0,00%   | 0,00%    | 39,51%   | 0,00%   | 7,94%   |
|        | 321 - 360 | 0,00%        | 0,00%   | 0,00%    | 40,53%   | 0,00%   | 8,14%   |
|        | 361+      | 0,00%        | 0,00%   | 0,00%    | 3,55%    | 0,00%   | 0,71%   |
| Gesamt |           | 100,00%      | 100,00% | 100,00%  | 100,00%  | 100,00% | 100,00% |

## A7. Anzahl der Zeilen der Beschreibungen

Anzahl

|        |   | Suchmaschine |              |            |            |            | Alle SM      |
|--------|---|--------------|--------------|------------|------------|------------|--------------|
|        |   | Ask          | Google       | MSN/Live   | Seekport   | Yahoo      |              |
|        | 1 | 184          | 130          | 131        | 23         | 39         | 507          |
|        | 2 | 826          | 875          | 835        | 27         | 898        | 3.461        |
|        | 3 | 0            | 1            | 5          | 134        | 1          | 141          |
|        | 4 | 0            | 0            | 0          | 795        | 0          | 795          |
|        | 5 | 0            | 0            | 0          | 8          | 0          | 8            |
| Gesamt |   | <b>1.010</b> | <b>1.006</b> | <b>971</b> | <b>987</b> | <b>938</b> | <b>4.912</b> |

Anteile

|        |   | Suchmaschine |         |          |          |         | Alle SM |
|--------|---|--------------|---------|----------|----------|---------|---------|
|        |   | Ask          | Google  | MSN/Live | Seekport | Yahoo   |         |
|        | 1 | 18,22%       | 12,92%  | 13,49%   | 2,33%    | 4,16%   | 10,32%  |
|        | 2 | 81,78%       | 86,98%  | 85,99%   | 2,74%    | 95,74%  | 70,46%  |
|        | 3 | 0,00%        | 0,10%   | 0,51%    | 13,58%   | 0,11%   | 2,87%   |
|        | 4 | 0,00%        | 0,00%   | 0,00%    | 80,55%   | 0,00%   | 16,18%  |
|        | 5 | 0,00%        | 0,00%   | 0,00%    | 0,81%    | 0,00%   | 0,16%   |
| Gesamt |   | 100,00%      | 100,00% | 100,00%  | 100,00%  | 100,00% | 100,00% |

Für die Berechnung wurde eine durchschnittliche Länge von 95 Zeichen pro Zeile zugrundegelegt.

## A8. Anteile Sprachtypen je Suchmaschine

In absoluten Zahlen

| <b>Sprachtyp</b>          | <b>Ask</b>  | <b>Google</b> | <b>MSN/Live</b> | <b>Seekport</b> | <b>Yahoo</b> | <b>Alle SM</b> |
|---------------------------|-------------|---------------|-----------------|-----------------|--------------|----------------|
| Vollständiger Satz        | 262         | 344           | 356             | 22              | 115          | <b>1099</b>    |
| Ausschnitt aus einem Satz | 639         | 547           | 526             | 707             | 713          | <b>3132</b>    |
| Einzelne Wörter           | 93          | 102           | 74              | 249             | 98           | <b>616</b>     |
| Fremdsprache              | 16          | 13            | 15              | 9               | 12           | <b>65</b>      |
| <b>Summe</b>              | <b>1010</b> | <b>1006</b>   | <b>971</b>      | <b>987</b>      | <b>938</b>   | <b>4912</b>    |

Nicht berücksichtigt: TB ohne Inhalt

|   |    |    |    |    |            |
|---|----|----|----|----|------------|
| 9 | 14 | 11 | 15 | 66 | <b>115</b> |
|---|----|----|----|----|------------|

Anteile

| <b>Sprachtyp</b>   | <b>Ask</b>     | <b>Google</b>  | <b>MSN/Live</b> | <b>Seekport</b> | <b>Yahoo</b>   | <b>Alle SM</b> |
|--------------------|----------------|----------------|-----------------|-----------------|----------------|----------------|
| Vollständiger Satz | 25,94%         | 34,19%         | 36,66%          | 2,23%           | 12,26%         | 22,37%         |
| Ausschn. aus Satz  | 63,27%         | 54,37%         | 54,17%          | 71,63%          | 76,01%         | 63,76%         |
| Einzelne Wörter    | 9,21%          | 10,14%         | 7,62%           | 25,23%          | 10,45%         | 12,54%         |
| Fremdsprache       | 1,58%          | 1,29%          | 1,54%           | 0,91%           | 1,28%          | 1,32%          |
| <b>Summe</b>       | <b>100,00%</b> | <b>100,00%</b> | <b>100,00%</b>  | <b>100,00%</b>  | <b>100,00%</b> | <b>100,00%</b> |
| <b>N</b>           | <b>1010</b>    | <b>1006</b>    | <b>971</b>      | <b>987</b>      | <b>938</b>     | <b>4912</b>    |

## A9. Anteile Sprachtypen je Beschreibungsquelle

in Absoluten Zahlen

| Sprachtyp                 | KWIC | Metadaten | ODP | Sonstiges |
|---------------------------|------|-----------|-----|-----------|
| Vollständiger Satz        | 284  | 585       | 181 | 49        |
| Ausschnitt aus einem Satz | 1792 | 1114      | 71  | 155       |
| Einzelne Wörter           | 451  | 116       | 4   | 45        |
| Fremdsprache              | 29   | 24        | 3   | 9         |
|                           | 2556 | 1839      | 259 | 258       |

Anteile

| Sprachtyp          | KWIC    | Metadaten | ODP     | Sonstiges |
|--------------------|---------|-----------|---------|-----------|
| Vollständiger Satz | 11,11%  | 31,81%    | 69,88%  | 18,99%    |
| Ausschn. aus Satz  | 70,11%  | 60,58%    | 27,41%  | 60,08%    |
| Einzelne Wörter    | 17,59%  | 6,34%     | 1,54%   | 17,44%    |
| Fremdsprache       | 1,13%   | 1,31%     | 1,16%   | 3,49%     |
| Summe              | 100,00% | 100,00%   | 100,00% | 100,00%   |
| N                  | 2556    | 1839      | 259     | 258       |

## A10. Anteile an Wikipedia Treffer

|                   | Ask | Google | MSN/ Live | Seekport | Yahoo |
|-------------------|-----|--------|-----------|----------|-------|
| Treffer           | 65  | 55     | 13        | 36       | 54    |
| Suchanfragen      | 62  | 49     | 12        | 36       | 44    |
| davon > 1 Treffer | 3   | 6      | 2         | 0        | 10    |

## A11. Bewertungen und Länge der Beschreibungen

In Wörter

| Anzahl_der_Wörter (klassiert) | Mittelwert | N     |
|-------------------------------|------------|-------|
| <= 5                          | 4,3529     | 102   |
| 6 - 10                        | 3,2282     | 298   |
| 11 - 15                       | 2,9779     | 588   |
| 16 - 20                       | 2,9583     | 1.414 |
| 21 - 25                       | 3,0892     | 1.144 |
| 26 - 30                       | 3,2364     | 275   |
| 31 - 35                       | 3,4876     | 121   |
| 36 - 40                       | 3,5266     | 207   |
| 41 - 45                       | 3,5187     | 214   |
| 46 - 50                       | 3,3977     | 171   |
| 51+                           | 3,5400     | 100   |
| Insgesamt                     | 3,1515     | 4.634 |

In Zeichen:

| Anzahl_der_Zeichen (klassiert) | Mittelwert | N     |
|--------------------------------|------------|-------|
| <= 40,00                       | 4,3875     | 80    |
| 41,00 - 80,00                  | 3,2600     | 250   |
| 81,00 - 120,00                 | 3,0079     | 507   |
| 121,00 - 160,00                | 3,0340     | 2.174 |
| 161,00 - 200,00                | 2,9865     | 743   |
| 201,00 - 240,00                | 3,6774     | 31    |
| 241,00 - 280,00                | 3,3506     | 77    |
| 281,00 - 320,00                | 3,5014     | 367   |
| 321,00 - 360,00                | 3,5499     | 371   |
| 361,00+                        | 3,6471     | 34    |
| Insgesamt                      | 3,1515     | 4.634 |

#### Eidesstattliche Versicherung

Ich versichere, die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst und keine anderen Quellen und Hilfsmittel als die angegebenen benutzt zu haben. Die aus anderen Werken wörtlich entnommenen Stellen, oder auch dem Sinn nach entlehnten Passagen, sind durch Quellenangabe kenntlich gemacht.

Hamburg, 12. August 2008

Lukas Tittmann