

Ilmenauer Beiträge zur Wirtschaftsinformatik

Herausgegeben von U. Bankhofer, V. Nissen
D. Stelzer und S. Straßburger

Daniela Büttner, Bernd Markscheffel

**Ein Vergleich ausgewählter
Desktop-Suchmaschinen**

Arbeitsbericht Nr. 2011-02, April 2011



Technische Universität Ilmenau
Fakultät für Wirtschaftswissenschaften
Institut für Wirtschaftsinformatik

Autor: Daniela Büttner, Bernd Markscheffel

Titel: Ein Vergleich ausgewählter Desktop-Suchmaschinen

Ilmenauer Beiträge zur Wirtschaftsinformatik Nr. 2011-02, Technische Universität Ilmenau, 2011

ISSN 1861-9223

ISBN 978-3-938940-33-4

urn:nbn:de:gbv:ilm1-2011200083

© 2011 Institut für Wirtschaftsinformatik, TU Ilmenau

Anschrift: Technische Universität Ilmenau, Fakultät für Wirtschaftswissenschaften,
Institut für Wirtschaftsinformatik, PF 100565, D-98684 Ilmenau.
<http://www.tu-ilmenau.de/wid/forschung/ilmenauer-beitraege-zur-wirtschaftsinformatik>

Gliederung

Gliederung	ii
Abbildungsverzeichnis	iv
Tabellenverzeichnis	iv
Abkürzungsverzeichnis	vi
1 Einleitung	1
2 Definition und Methodik	3
3 Kriterienkatalog	5
3.1 Software-Ebene	5
3.1.1 Vielseitigkeit und Einschränkungen	6
3.1.2 Effizienz	6
3.1.3 Usability	7
3.2 Suchmaschinen-Ebene	7
3.2.1 Qualität der Suchfunktion	7
3.2.2 Qualität der Suchresultate	10
3.3 Desktop-Suchmaschinen-Ebene	11
3.3.1 Indexierung	12
3.3.2 Anpassungsfähigkeit	13
3.3.3 Sicherheit	14
3.3.4 Darstellung der Suchergebnisse	14
4 Vorstellung der Desktop-Suchmaschinen	16
4.1 Windows 7-Suche	16
4.2 Copernic Desktop Search	16
4.3 Google Desktop	17

4.4	Hulbee Desktop.....	17
4.5	Xfriend personal Desktop Search	18
4.6	Archivarius 3000.....	18
4.7	Find and run robot (FARR).....	19
5	Testumgebung	20
5.1	Technische Daten.....	20
5.2	Testkollektion	21
6	Suchmaschinen-Vergleich.....	23
6.1	Vielseitigkeit & Einschränkungen	23
6.2	Effizienz.....	25
6.3	Usability.....	29
6.4	Sprachliche Analyse.....	31
6.5	Qualität der Rechtschreibprüfung.....	35
6.6	Analyse bestimmter Suchfunktionen	36
6.7	Erweiterte Suchmöglichkeiten	41
6.8	Retrievalmaße	42
6.9	Indexierung	46
6.10	Anpassungsfähigkeit	49
6.11	Vertraulichkeit	53
6.12	Darstellung der Suchergebnisse.....	54
7	Auswertung	58
7.1	Software-Ebene.....	59
7.2	Suchmaschinen-Ebene	60
7.3	Desktop-Suchmaschinen-Ebene.....	61
7.4	Gesamtwertung	62
8	Zusammenfassung und Ausblick.....	63

Literaturverzeichnis	66
Anhang	69

Abbildungsverzeichnis

Abb. 1: Relevanzbewertung von Suchergebnissen.....	11
Abb. 2: Größe von Installationsdatei und Programmordner.....	25
Abb. 3: Auslastung der CPU	27
Abb. 4: Auslastung des Arbeitsspeichers	27
Abb. 5: Relativer Recall	45
Abb. 6: Indexierungsadauer.....	48
Abb. 7: Indexgröße	49

Tabellenverzeichnis

Tab. 1: Komponenten des Testsystems	20
Tab. 2: Anwendungssoftware des Testsystems	21
Tab. 3: Dateiformate der Testkollektion	22
Tab. 4: Vielseitigkeit und Einschränkungen.....	23
Tab. 5: Stemming	32
Tab. 6: Unterscheidung von Singular und Plural	33
Tab. 7: Soundex.....	34
Tab. 8: Rechtschreibprüfung und –korrektur	35
Tab. 9: Boolesche Operatoren und Abstandsoperator.....	36
Tab. 10: Standardverknüpfung	37
Tab. 11: Trunkierungsoperatoren	38

Tab. 12: Inkrementelle Suche	40
Tab. 13: Erweiterte Suchmöglichkeiten	42
Tab. 14: Gesamttrefferzahl	42
Tab. 15: Precision	44
Tab. 16: Relevante Ergebnisse mit Ersatzwerten	45
Tab. 17: Beeinflussung der Indexierung durch den Benutzer	46
Tab. 18: Darstellungsformen	50
Tab. 19: Spezielle Anwendungen	51
Tab. 20: Unterstützung gelöschter und gepackter Dateien	53
Tab. 21: Darstellung der Suchergebnisse	56
Tab. 22: Bewertung der Testergebnisse auf Software-Ebene	59
Tab. 23: Bewertung der Testergebnisse auf Suchmaschinen-Ebene	60
Tab. 24: Bewertung der Testergebnisse auf Desktop-Suchmaschinen-Ebene	61
Tab. 25: Zusammenfassung der Ergebnisse aller Ebenen	62

Abkürzungsverzeichnis

CPU	Central Processing Unit
DDR	Double Data Rate
DIMM	Dual Inline Memory Module
GB	Gigabyte
IT	Information Technology
MB	Megabyte
PC	Personal Computer
RAM	Random-Access-Memory
SATA	Serial Advanced Technology Attachment

Zusammenfassung: Die ambivalente Entwicklung auf dem Festplattenspeichermarkt - sinkende Preise bei gleichzeitig steigender Speicherkapazität - hat die Speicherung großer heterogener Datenmengen für den Nutzer von PCs wesentlich vereinfacht und damit die Notwendigkeit des Einsatzes effektiver Retrievalmechanismen in Form von Desktopsuchmaschinen zusätzlich motiviert.

Da mittlerweile eine breite Auswahl an solchen Suchmaschinen für den PC angeboten wird und nur wenige wissenschaftliche Arbeiten zum Thema vorliegen, sollen in diesem Forschungsbericht ausgewählte aktuelle Desktop-Suchmaschinen in einem praktischen Test verglichen werden.

Für diesen Vergleich werden zunächst geeignete Kriterien aus bisherigen Arbeiten zu Desktop-Suchmaschinen, sowie aus Untersuchungen, die den Vergleich von Suchmaschinen und Software im Allgemeinen betrachten, abgeleitet.

Nach der Komposition eines geeigneten Kriterienkataloges folgt die eigentliche Testphase. Für jede Suchmaschine und jedes Leistungsmerkmal werden in einer homogenen Testumgebung die entsprechenden Untersuchungen durchgeführt und dokumentiert. Im Anschluss an die Evaluation wird dann der Vergleich im Hinblick auf die verschiedenen Kriterien durchgeführt. So kann transparent gemacht werden, in welchen Bereichen die einzelnen Programme Stärken bzw. Schwächen aufweisen.

Der vorgestellte Kriterienkatalog kann zudem als Basis für nachfolgende Studien zur Leistungsmessung von Desktop-Suchmaschinen verwendet werden. Da bezüglich der Entwicklungen des Speichermedienmarktes nicht von einer Trendwende auszugehen ist, wird die Bedeutung der Desktop-Suchmaschinen für PC-Anwender in Zukunft noch weiter steigen.

Schlüsselworte: Desktop-Suchmaschinen, Suchmaschinen, Software-Test, Vergleich

1 Einleitung

Betrachtet man die Entwicklung der Speicherkapazitäten von Festplatten innerhalb der letzten Jahre, so lässt sich ein enormes Wachstum feststellen. Waren im Jahr 2002 noch Größen im Bereich von 250 Gigabyte ein Novum für interne Festplatten,¹ so brachte der Hersteller Western Digital 2009 das erste Speichermedium mit der nahezu zehnfachen Kapazität von zwei Terabyte auf den Markt.² Im Gegensatz zu diesem Trend lässt sich bei der Preisentwicklung genau Gegenteiliges beobachten: Seit 2005 sanken die Preise für Festplatten um 70%.³

Diese Entwicklungen, sowie die Entstehung und Verwendung von immer neuen Dateiformaten, vor allem im Multimedienbereich, haben dazu geführt, dass Benutzer heute eine beträchtliche Menge unterschiedlichster Daten auf ihren Rechnern abspeichern. Schwierigkeiten treten meist auf, wenn es um das Wiederfinden der einmal gespeicherten Informationen geht [Cole05, 14]. Obgleich gewissenhaft angelegte Ordnerstrukturen den Suchaufwand gering halten können, stoßen diese irgendwann an ihre Grenzen und andere Anwendungen müssen weiterhelfen [BlSc07, 1].

Microsoft bietet deshalb zwar eine in das Betriebssystem Windows integrierte Suchfunktion an, jedoch hat sich das Leistungsvermögen als unzureichend herausgestellt, um allen Anforderungen gerecht zu werden [Cole05, 14]. Diese Lücke haben Anbieter von Internet-Suchmaschinen wie Google und Yahoo, sowie andere Firmen, z.B. Copernic Technologies, gleichermaßen erkannt und haben für den Einsatz am eigenen PC optimierte Suchprogramme, sogenannte Desktop-Suchmaschinen, entwickelt.

Betrachtet man die Tatsache, dass eine effektive Nutzung des Internets ohne Suchmaschinen nahezu unmöglich wäre [Woll05, 1], so erscheint eine ähnliche Entwicklung im Hinblick auf den privaten Rechner nicht unrealistisch, wenn die eingangs beschriebenen Trends der letzten Jahre weiter voranschreiten. Für den Benutzer wird langfristig gesehen

¹ Vgl. <http://www.golem.de/0211/22824.html>. 2002, Abruf: 2009-10-28

² Vgl. http://www.pcwelt.de/start/computer/festplatte_storage/news/191873/erste_2_terabyte_festplatte_gesichtet_update_jetzt_offiziell/, Abruf: 2009-10-28.

³ Vgl. http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/pm/2009/07/PD09__287__614.psm1, Abruf : 2009-10-28

kein Weg mehr an der Verwendung einer Desktop-Suchmaschine vorbeiführen, wenn er gesuchte Informationen auch effizient finden möchte.

Ziel dieses Arbeitsberichtes ist ein Vergleich ausgewählter Desktop-Suchmaschinen.

Die Programme werden hierfür in einer empirischen Untersuchung auf Bewertungskriterien getestet, die aus einer Analyse von unterschiedlichen Softwarevergleichen abgeleitet wurden.

Nach der Einleitung in Kapitel 1 wird im nächsten Abschnitt der Begriff der Desktop-Suchmaschine diskutiert sowie die methodische Basis des Vergleichs erläutert. In Kapitel 3 folgt die Ableitung des Kriterienkatalogs, bevor in Kapitel 4 die untersuchten Desktop-Suchmaschinen vorgestellt werden. Die dem Vergleich zugrunde liegende Testumgebung wird im folgenden Kapitel beschrieben. Daran anschließend wird in Kapitel 6 der konkrete Vergleich der Desktop-Suchmaschinen vorgenommen. In Kapitel 7 erfolgt die Auswertung der Ergebnisse des Vergleichs bevor in Kapitel 8 die wesentlichen Ergebnisse der Arbeit zusammengefasst und einer kritischen Würdigung unterzogen werden.

2 Definition und Methodik

Unter einer Desktop-Suchmaschine versteht man eine Softwareanwendung, welche Dateien auf einem PC indexiert und sucht [LSSW, 110]. Sie ist in der Lage die gesuchten Dateien anhand von Benutzeranfragen, die Stichworte, Dateitypen oder bestimmte Verzeichnisse o.ä. beinhalten, zu finden.

Bei der vorliegenden Arbeit handelt es sich um eine empirische Untersuchung von Desktop-Suchmaschinen. Grundlage ist eine Literaturstudie [WeWa02], in der bisherige Benchmarking-Ansätze von Desktop-Suchmaschinen, sowie vergleichende Untersuchungen von Suchmaschinen und Software analysiert werden. Ziel hierbei ist es, zunächst die zu untersuchenden Desktop-Suchmaschinen, sowie geeignete Bewertungskriterien für den Suchmaschinenvergleich zu ermitteln.

Ausgangspunkt für die Liste der zu untersuchenden Tools ist die Benchmark-Studie von Noda und Helwig aus dem Jahr 2005 [NoHe05]. In dieser Studie bewertete und heute noch relevante Suchmaschinen werden in ihrer aktuellen Version betrachtet. Die Liste wird ergänzt um ausgewählte Software-Anwendungen, die seitdem neu auf den Markt gekommen sind. Zur Identifizierung dieser dient die Recherche mittels Web-Suchmaschinen. Es werden lediglich entgeltfreie und solche Anwendungen getestet, für die eine gebührenfreie Testversion bereitgestellt wird.

Im Hinblick auf das Ziel der empirischen Untersuchung, ein allgemeines Ranking der untersuchten Desktop-Suchmaschinen auf Basis der Bewertung verschiedener Einzelkriterien zu erstellen, müssen im zweiten Schritt diese Bewertungskriterien festgelegt werden. Um den Kriterienkatalog der Evaluation in überschaubarem Umfang zu halten, findet eine Beschränkung auf die aus Nutzersicht relevanten Aspekte statt. Bei der Auswahl dieser wird u.a. die Arbeit von Burmeister und Schirrmann mit einbezogen [BuSc06, 6 ff.], die im Vorfeld ihrer Desktop-Suchmaschinen-Analyse Studierende aus dem Seminar „Angewandte Informatik“ nach deren Prioritäten im Hinblick auf die vorher zusammengestellten Kriterien befragten.

Als Rahmen für den Aufbau des Katalogs eignet sich die Arbeit von Kirchner und Jung [KiJu01, 4]. Ihr Konzept zur Bewertung von UML-Werkzeugen auf drei verschiedenen Abstraktionsebenen lässt sich gut auf den Vergleich von Desktop-Suchmaschinen übertra-

gen. Dazu werden die hier verwendeten Bewertungskriterien der einzelnen Ebenen auf Software-Ebene - anstelle der Kriterien für Modellierungswerkzeuge und UML-Werkzeuge - um die für den betrachteten Kontext relevanten Kriterien für Suchmaschinen und Desktop-Suchmaschinen ergänzt. Angelehnt an Kirchner und Jung wird die erste und allgemeinste Bewertungsebene in dieser Arbeit die Software-Ebene sein, da natürlich eine Desktop-Suchmaschine grundlegend betrachtet eine Software zur Zielerreichung in einem bestimmten Einsatzbereich ist. Nutzerrelevante Qualitätskriterien, die zur Bewertung jeglicher Art von Software-Anwendung dienen, sind beispielsweise Usability und die Vielfalt an angebotenen Versionen [KiJu01, 6 ff.].

Spezialisiert man eine Ebene tiefer, kann die Software „Desktop-Suchmaschine“, wie der Name schon sagt, allgemein als „Suchmaschine“ klassifiziert werden. Für Suchmaschinen gelten spezielle Bewertungskriterien, die in der Arbeit ebenfalls untersucht werden sollen. Wichtig sind die Qualität des Index, der Suchfunktion und der Suchresultate, wobei bei Letzterer insbesondere die Maße Precision und Recall von Bedeutung sind [LeHö07, 3].

Auf der untersten Klassifizierungsebene werden Bewertungskriterien ermittelt, die einzig für die Betrachtung von Desktop-Suchmaschinen relevant sind und nicht für andere Arten von Software oder Suchmaschinen im Allgemeinen anwendbar sind. Hierzu zählen u.a. Messungen zur Indexierung der auf dem Rechner befindlichen Dateien, beispielsweise die Dauer der Erstindexierung, Größe des Index u.v.a.m. [NoHe05, 13 f.].

3 Kriterienkatalog

Der im folgenden Abschnitt vorgestellte Kriterienkatalog zur Bewertung von Desktop-Suchmaschinen soll dazu dienen, das Hauptziel der Arbeit, einen Vergleich ausgewählter Desktop-Suchmaschinen, zu erreichen. Darüber hinaus soll mit diesem Katalog eine Orientierungsvorlage für nachfolgende Qualitätsforschungen im Bereich der Desktop-Suchmaschinen geschaffen werden, denn auch zukünftig werden empirische Untersuchungen im Hinblick auf die Beschreibung und Bewertung der Leistungsfähigkeit von Desktop-Suchmaschinen notwendig sein, da sich der Markt für Desktop-Suchmaschinen durch eine hohe Dynamik auszeichnet, wo etablierte Anbieter kontinuierlich Neuversionen ihrer Software-Angebote auf den Markt bringen und Marktteilnehmer ständig wechseln. Betrachtet man in diesem Zusammenhang die Studie von Noda und Helwig aus dem Jahr 2005 [NoHe05, 5], so lässt sich feststellen, dass die bestplatzierte Suchmaschine Copernic Desktop Search, Version 1.5 Beta, heute bereits in der Version 3.4.0 erhältlich ist.⁴ Yahoo! Desktop Search fand man beispielweise zum Zeitpunkt der Untersuchung noch unter dem Namen X1 Desktop Search⁵. Mittlerweile ist diese wie auch andere Teilnehmer, z.B. Hot Bot, komplett vom Markt zurückgezogen worden. Andere, wie Hulbee, sind neu hinzugekommen.⁶

In den nachfolgenden Abschnitten werden die Bewertungskriterien für die im vorigen Kapitel vorgestellten Abstraktionsebenen erläutert.

3.1 Software-Ebene

Den Rahmen für die Bewertungskriterien auf dieser Ebene bildet die Arbeit von Kirchner und Jung [KiJu01], da die hier genannten Kriterien zur Beurteilung von UML-Werkzeugen auf Software-Ebene zu einem großen Teil auch für Desktop-Suchmaschinen sinnvoll sind.

⁴ Vgl. <http://www.copernic.com/en/products/desktop-search/home/download.html>, Abruf: 2011-02-01.

⁵ Vgl. <http://us.config.toolbar.yahoo.com/yds>, Abruf: 2010-06-20.

⁶ Vgl. [http://company.hulbee.com/news/?tx_ttnews\[pointer\]=1&tx_ttnews\[tt_news\]=28&tx_ttnews\[backPid\]=28&cHash=f55fb0161f](http://company.hulbee.com/news/?tx_ttnews[pointer]=1&tx_ttnews[tt_news]=28&tx_ttnews[backPid]=28&cHash=f55fb0161f), Abruf:2010-06-20.

3.1.1 Vielseitigkeit und Einschränkungen

Da komplexe programmspezifische Sachverhalte in der Muttersprache einfacher aufgenommen werden können als in einer Fremdsprache und somit die Einarbeitung in das Programm wesentlich vereinfacht wird, soll ermittelt werden, ob die Desktop-Suchmaschine in einer deutschsprachigen Version verfügbar ist [BuSc06, 7].

Wie erwähnt, hat Microsoft mit dem Betriebssystem Windows zwar nach wie vor den eindeutig größten Marktanteil mit 91,3% im Mai 2010, jedoch kann ein leichter Abwärtstrend über die vergangenen Monate festgestellt werden, wohingegen vor allem die Zahl der Mac-Nutzer zu steigen scheint.⁷ Folglich soll untersucht werden, ob die Anbieter der Desktop-Suchmaschinen lediglich eine mit Windows kompatible Version bereitstellen, oder ob auch andere Betriebssysteme in einer eigenen Programmversion berücksichtigt werden.

Weiterhin wird geprüft, ob das Funktionsspektrum der Desktop-Suchmaschinen durch Add-Ons erweitert werden kann. Da die Programme aus Sicht des Standardnutzers betrachtet werden, soll hierbei nur auf kostenlose Erweiterungsangebote eingegangen werden und nicht auf evtl. angebotene kostenpflichtige Professional- oder Unternehmensversionen.

Falls es sich bei der zu testenden Desktop-Suchmaschine um lizenzierte Software handelt, aber eine kostenfreie Testversion verfügbar ist, sollen eventuell vorhandene Einschränkungen dieser Version herausgefunden werden. Als Quellen dienen die Internetseiten der Anbieter, sowie auftretende Hinweise während des Installationsvorgangs und im Programmbetrieb. Falls die Einschränkungen untersuchte Kriterien betreffen, können sie sich negativ auf die Vergleichbarkeit der Desktop-Suchmaschinen auswirken.

3.1.2 Effizienz

Unter Effizienz soll der sparsame Verbrauch von Betriebsmitteln verstanden werden, welches aus Nutzersicht ein wichtiges Kriterium zur Bewertung von Software-Anwendungen darstellt [Endr03, 21]. In diesem Zusammenhang wird zunächst die Größe der herunterzuladenden Installationsdatei, sowie die Größe des durch die Installation erstellten Programmordners für die einzelnen Desktop-Suchmaschinen ermittelt und verglichen [NoHe05, 11; BuSc06, 8].

⁷ Vgl. <http://marketshare.hitslink.com/operating-system-market-share.aspx?qprid=9>, Abruf: 2010-06-20.

Außerdem wird im Hinblick auf die technische Effizienz die Speicherauslastung ermittelt, denn um dem Benutzer einen ungestörten Arbeitsablauf zu garantieren, sollte der Speicherbedarf der Desktop-Suchmaschine so gering wie möglich sein. Untersucht wird die Auslastung von CPU und Arbeitsspeicher während des Programmbetriebs, was bedeutet, dass die Suchmaschine geöffnet aber nicht in Gebrauch genommen wird. Weiterhin wird die Auslastung während einer Suchanfrage und während der Indexierung verglichen [BuSc06, 8].

3.1.3 Usability

In den von Endres vorgestellten Qualitätskriterien für Software wird unter anderem die Usability als ein wesentliches Kriterium für die Bewertung vorgestellt [Endr03, 21]. Noda und Helwig verwenden Usability ebenfalls als Kriterium für den Vergleich von Desktop-Suchmaschinen [NoHe05, 3]. Zur Usability zählt man nach ISO 9126 die Attraktivität der Gestaltung der Benutzeroberfläche, die Lernbarkeit, die Verständlichkeit sowie die technische Bedienbarkeit [Folm05, 65]. Im vorliegenden Arbeitsbericht wird beschrieben werden, was während der Testphase im Hinblick auf diese Gesichtspunkte im Umgang mit den jeweiligen Suchmaschinen positiv bzw. negativ aufgefallen ist. Im Kontext Usability wird zudem die Konfigurierbarkeit der Benutzeroberfläche untersucht, also in wieweit für den Anwender Möglichkeiten bestehen, eigenständig Anpassungen zur Gestaltung der Interaktion mit dem System vorzunehmen.

3.2 Suchmaschinen-Ebene

Die Bewertung der Desktop-Suchmaschinen auf Suchmaschinen-Ebene erfolgt in Anlehnung an Lewandowskis und Höchstötters Artikel „Qualitätsmessung bei Suchmaschinen“, wobei die Evaluationsbereiche „Qualität der Suchfunktion“ und „Qualität der Suchresultate“ als Oberkategorien für die einzelnen Untersuchungskriterien übernommen werden [LeHö07, 3].

3.2.1 Qualität der Suchfunktion

Suchmaschinen verfügen über spezielle Funktionen, die dem Benutzer die Suche ermöglichen bzw. erleichtern, was wiederum Einfluss auf die Qualität der Ergebnisse hat [LeHö07, 3]. Mittels der nachfolgenden Kriterien sollen diese Funktionen bewertet werden.

Sprachliche Analyse

In der Untersuchung des Umgangs mit sprachlichen Besonderheiten soll zunächst herausgefunden werden, ob die Suchmaschinen Stemming unterstützen. Unter Stemming versteht man die Reduktion eines Wortes auf den Wortstamm [BiCe09, 43].⁸ Durch diesen kommt in der Regel die Grundbedeutung des Wortes zum Ausdruck, die durch ergänzende Wortbildungsmechanismen wie Prä-, Post- oder Zirkumfixe nicht verändert wird.

Weiterhin wird im Test untersucht, ob die Suchmaschinen zwischen Singular und Plural unterscheiden, d.h. ob bei Eingabe der Singularform auch Pluralformen ausgegeben werden und umgekehrt.

Darüber hinaus soll die Auswirkung von Groß- bzw. Kleinschreibung des Suchbegriffs auf die Suchergebnisse getestet werden [BiCe09, 210]. Wichtig wird eine Unterscheidung vor allem, wenn durch die andere Schreibweise eine Bedeutungsänderung entsteht, wie das z.B. bei „it“ und „IT“ der Fall wäre.

Eine Schwierigkeit im Bereich Information Retrieval stellen seit jeher Namen und ihre möglichen unterschiedlichen Schreibweisen dar [Stoc07, 307]. Die Mehrdeutigkeit bei der Repräsentation hat unter anderem zur Entwicklung des sogenannten Soundex-Algorithmus geführt, mit dessen Hilfe gleich gesprochene Lexeme (Homophone), die aber unterschiedlich geschrieben werden, einheitlich codiert werden. Unterstützt eine Suchmaschine diesen Algorithmus werden beispielsweise bei der Suchanfrage nach dem Namen „Mayer“ auch identisch ausgesprochene Formen, die sich jedoch in der Schreibweise unterscheiden, gefunden, wie z.B. „Maier“ oder „Meier“.

Qualität der Rechtschreibprüfung

In engem Zusammenhang mit dem letzten Abschnitt kann die Rechtschreibprüfung gesehen werden, denn wenn Rechtschreibfehler von den Suchmaschinen entdeckt werden, ist das in der Regel auch auf die Unterstützung des Soundex-Algorithmus zurückzuführen [Stoc07, 306 f.]. Die am häufigsten auftretenden Fehlertypen sind die Verwendung eines falschen Buchstabens bei gleicher Wortlänge, das Auslassen eines Buchstabens, das Einfügen eines zusätzlichen Buchstabens und Buchstabendreher. Im Suchmaschinentest wird die Reaktion der einzelnen Programme auf diese Fehler untersucht.

⁸ Bsp.: Der Term „spielerisch“ ist eine vom Lexem „Spiel“ abgeleitete Form.

Bestimmte Suchfunktionen

Neben der rein auf die Sprache bezogenen Funktionalität verfügen Suchmaschinen über zahlreiche andere Funktionen, die den Benutzer bei der Suchfragenformulierung unterstützen sollen.

Eine dieser spezifischen Funktionen ist das Unterstützen von booleschen Operatoren, welche der Verknüpfung von Suchtermen dienen [CNSe04, 498]. Durch die Verwendung von AND werden Ergebnisse geliefert, die alle Suchbegriffe enthalten. Verbindet man zwei Suchterme mit OR, so enthalten die Ergebnisse entweder den einen und/oder den anderen Suchterm. Der Operator NOT eignet sich, um bestimmte Begriffe von der Suchergebnismenge auszuschließen. Suchmaschinen, die boolesche Operatoren unterstützen, bieten den Vorteil, dass Suchanfragen präziser formuliert werden können und somit das Suchergebnis genauer wird. Weiterhin wird das Vorhandensein von Abstandsoperatoren (bspw. NEAR) [Sünk09, 7] geprüft, durch die Dokumente gefunden werden, in denen die Suchbegriffe als Kollokation auftreten. Weiterhin ist zu klären, welcher der genannten Operatoren als Standardverknüpfung verwendet wird.

Eine wichtige Rolle für die Suchanfragenformulierung spielen die Trunkierungsoperatoren. Sie bieten dem Benutzer die Möglichkeit, Varianten eines Wortes (bspw. Flexionen) zu finden, indem es nicht vollständig ausgeschrieben wird, sondern ein Platzhalter, der Trunkierungsoperator, verwendet wird [Koch07, 34]. Ein häufig benutztes Symbol ist das Asterisk (Unicode U+002A), welches beliebig viele Zeichen ersetzt [PiWo09, 363].

Eine erhebliche Erleichterung für den Benutzer bietet die Speicherung der Suchanfragen, da bei wiederholter Suche nach gleichen Begriffen die Zeit für eine erneute Eingabe eingespart werden kann [BuSc06, 14]. Zudem kann der Nutzer die genaue Formulierung der Anfragen erkennen, welche er evtl. bereits zu einem bestimmten Begriff gestellt hat. Es wird getestet, ob die Suchmaschinen diese Funktion unterstützen bzw. ob die gespeicherten Terme auch nach einem Neustart des Computers noch vorhanden sind.

Aus Gründen der Zeitersparnis und Vereinfachung ist die inkrementelle Suche sinnvoll, bei welcher bereits während der Eingabe des Suchbegriffs Ergebnisse angezeigt werden, die mit den schon eingetippten Buchstaben beginnen und tatsächlich in der Datenbasis vorhanden sind (ähnlich wie bei Google-Instant Search).⁹ Im Test soll ermittelt werden, ob die

⁹ Vgl. http://www.itfrontal.de/2004/12/google_jetzt_mi.html, Abruf: 2010-08-31.

Desktop-Suchmaschinen diese Funktion unterstützen, oder ob erst nach Bestätigung mit „Enter“ oder einem „Suche“-Button die Resultate angezeigt werden.

Weiterhin soll geprüft werden, welche Bereiche der Datei die Suchmaschine durchsucht, ob nur der Dateiname oder auch der Dateiinhalt berücksichtigt wird.

Erweiterte Suchmöglichkeiten

Web-Suchmaschinen bieten in der Regel erweiterte Such- bzw. Filtermöglichkeiten an, über die der Benutzer seine potenzielle Ergebnismenge gezielt einschränken kann.¹⁰

Zunächst wird getestet, ob die Programme die unter anderem in der „erweiterten Suche“ der Web-Suchmaschine Google vorhandenen Optionen anbieten, wonach Ergebnisse gefunden werden sollen, die beispielsweise „Alle“ eingegebenen Wörter beinhalten. Außerdem werden die Desktop-Suchmaschinen auf die Einschränkungsmöglichkeiten getestet, nur nach Dateien in einer bestimmten Sprache oder in einem bestimmten Format zu suchen. Weitere Kriterien sollen die Einschränkungen bzgl. des Erstellungs- und Veränderungsdatums, sowie auf die Position des Suchbegriffs, im Dateinamen oder im Inhalt sein.

3.2.2 Qualität der Suchresultate

Um im folgenden Vergleich die Qualität der Suchresultate bzw. die Retrievaleffektivität der einzelnen Suchmaschinen zu testen, werden neben der Ermittlung der Gesamttrefferrzahl klassische Retrievalmaße verwendet [LeHö07, 3].

Grundsätzlich gilt, dass sich ein Nutzer von der Anwendung einer Suchmaschine relevante Ergebnisse erhofft, was bedeutet, dass eine semantische Nähe zwischen der Suchanfrage und dem Inhalt einer Information vorhanden sein muss bzw. dass der Informationsgehalt dem der Frage entspricht [Mark10, 35]. Einen höheren Nutzen für den Anwender hat das gefundene Dokument, wenn neben Relevanz auch Pertinenz besteht, d.h. dass die gefundene Information auch dem Informationsbedarf des Nutzers entspricht.

Als Qualitätskriterien für die Retrievaleffektivität werden Precision und Recall verwendet. Mittels der Precision wird die Genauigkeit der Suche gemessen. Das Maß beschreibt das Verhältnis der gefundenen relevanten Dokumente (R_a) zur Anzahl aller gefundenen Dokumente (A) [Mark10, 35]. Das Maß für die Vollständigkeit des Retrievalergebnisses wird

¹⁰ http://www.google.de/advanced_search?hl=de, Abruf: 2010-06-22.

als Recall bezeichnet. Er kennzeichnet das Verhältnis zwischen den gefundenen relevanten Dokumenten (R_a) und der insgesamt im Datenbestand vorhandenen relevanten Dokumente (R). In der nachfolgenden Abbildung 1 sind die beschriebenen Dokumentenmengen veranschaulichend dargestellt.

Da eine genaue Aussage über die Anzahl aller relevante Dokumente zu einer Suchanfrage (in Abb. 1: die Menge R) und somit des gesamten Recall in realen Textkorpora, nur sehr schwer möglich ist, soll ein Hilfsverfahren verwendet werden, welches Lewandowski als „Pooling“ [Lewa07, 14] bezeichnet. Hierbei werden die durch das Maß Precision ermittelten relevanten Treffer aller Suchmaschinen in einem Pool zusammengefasst. Der relative Recall stellt dann den Anteil der jeweiligen Suchmaschine an den von allen Suchmaschinen ausgegebenen relevanten Treffern dar.

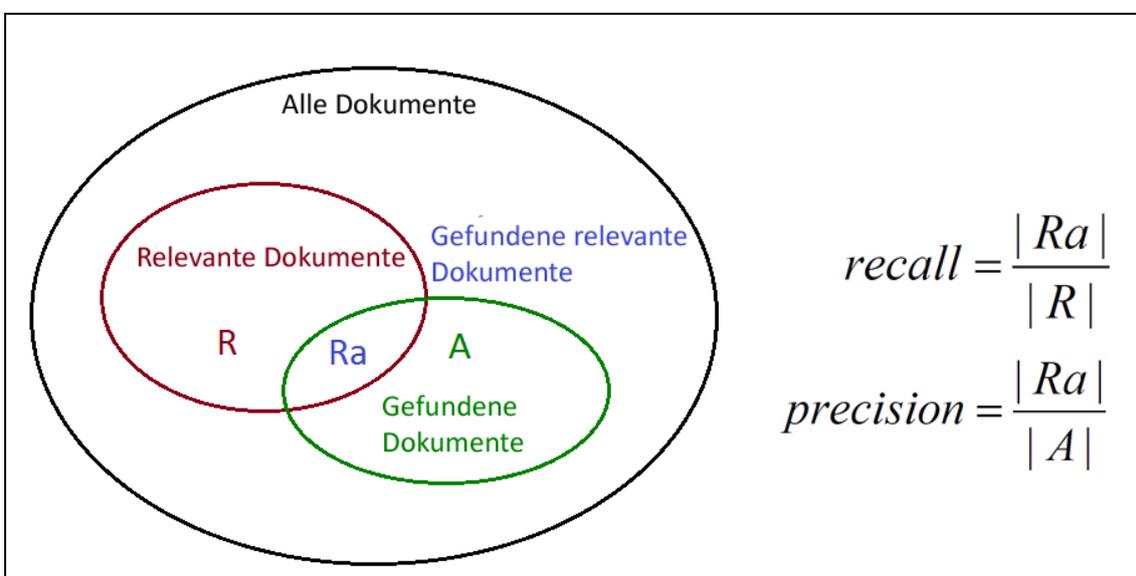


Abb. 1: Relevanzbewertung von Suchergebnissen¹¹

3.3 Desktop-Suchmaschinen-Ebene

In Ergänzung zu den bereits auf Suchmaschinen-Ebene behandelten Evaluationsbereichen „Qualität der Suchfunktion“ und „Qualität der Suchresultate“ sowie der auf Software-Ebene eingeordneten „Usability“, schlägt Lewandowski als weiteren Untersuchungsbereich die Qualität des Index vor [LeHö07, 3]. Da Benutzer von Desktop-Suchmaschinen im Regelfall, anders als bei den Web-Suchmaschinen, Einfluss auf die Indexierung nehmen

¹¹ In Anlehnung an [Mark10, 36 f.].

können und die meisten getesteten Kriterien im Suchmaschinen-Kontext keine Rolle spielen, werden die Aspekte der Indexierung auf der Ebene der Desktop-Suchmaschine erfasst. Weitere Kriterien, wie Sicherheitsaspekte, werden in Anlehnung an Noda und Helwig ebenfalls in den Katalog aufgenommen [NoHe05, 14].

3.3.1 Indexierung

Mit der Indexierung wird die Basis für schnelle Suchanfragen geschaffen [Cole05, 15]. Statt der kompletten Datei wird bei einer Suchanfrage ein vorher erstellter Index (in der Regel eine invertierte Datei) durchsucht, welcher u.a. Dateiname, -typ und Metadaten jeder Datei der Datenbasis enthält.

Da die Desktop-Suchmaschinen aus Benutzersicht getestet werden, soll zunächst untersucht werden, in wieweit der Anwender Einfluss auf den Indexierungsvorgang nehmen kann. So wird ermittelt, ob eine Auswahlmöglichkeit im Hinblick auf die zu indexierenden Dateiformate und Verzeichnisse getroffen werden kann und ob der Anwender den Zeitpunkt der Indexierung steuern kann. Hierdurch entstehen Vorteile für den Nutzer, wie etwa die Steuerbarkeit des Speicherbedarfs für die Indexdatei oder das Verlagern der systemressourcenintensiven Indexierung auf geeignete Zeiten.

Eine neu erstellte Datei kann erst nach der Aufnahme in den Index von der Suchmaschine gefunden werden. Aus diesem Grund soll die Aktualität des Index gemessen werden, also wie viel Zeit zwischen Erstellung einer Datei und dem erstmaligen Auffinden nach Formulierung einer passenden Suchanfrage liegt.

Als weiteres Gütekriterium für die Indexierung kann die Indexierungsdauer genannt werden. Da während dieses Vorgangs eine erhöhte Ressourcenauslastung besteht, ist es wichtig, diesen schnellstmöglich durchzuführen, um die Beeinträchtigung für den Benutzer gering zu halten. Gerade die Erstindexierung kann unter Umständen verhältnismäßig lange dauern. Bei Google Desktop ist die Rede von einer möglichen Erstindexierungsdauer von mehreren Tagen.¹² Im Suchmaschinenvergleich wird gemessen, wie lange die einzelnen Programme für die vollständige Indexierung der Testkollektion benötigen.

Außerdem wird die Größe der erstellten Indexdatei für jede Desktop-Suchmaschine ermittelt.

¹² Vgl. <http://desktop.google.de/support/bin/answer.py?hl=de&answer=12403>, Abruf: 2010-06-22

3.3.2 Anpassungsfähigkeit

Zunächst soll geprüft werden, wie gut Desktop-Suchmaschinen an unterschiedliche Verwendungsmöglichkeiten angepasst sind, indem untersucht wird, ob verschiedene Arten von Benutzeroberflächen dem Anwender einer Suchmaschine zur Verfügung stehen. So ist es möglich, dass es sich bei der Desktop-Suchmaschine um ein eigenständiges Programm handelt oder dass sie im Browser integriert ist [NoHe05, 12]. Des Weiteren kann die Sucheingabe beispielsweise über eine Deskbar (ein in die Taskleiste integriertes Suchfeld) getätigt werden.¹³ Ist dieses frei auf dem Desktop platzierbar, spricht man von einer Floating Deskbar. Als Schnellsuchfeld wird ein ebenso auf dem Desktop bewegliches Fenster bezeichnet, welches durch eine bestimmte Tastenkombination aufgerufen werden kann. Weiterhin besteht die Option, dass ein Sucheingabefeld in einer Seitenleiste angezeigt wird. Für die Windows 7-Suche befindet sich das Suchfeld im Startmenü und in den geöffneten Fenstern des Windows-Explorers. Ist neben der Desktop-Suchmaschine ebenfalls eine Web-Suchmaschine vom gleichen Anbieter verfügbar, so wird oft eine Toolbar für den Internetbrowser zur Installation angeboten.¹⁴ Da von den Desktop-Suchmaschinen, wie am Beispiel von Google Desktop erkennbar, oft mehrere Darstellungsformen angeboten werden, werden für jede Suchmaschine alle möglichen Varianten dokumentiert.

Im nächsten Punkt wird geprüft, welche Dateitypen, insbesondere auch Medien-Dateiformate unterstützt werden [BuSc06, 8].

Des Weiteren wird analysiert, ob die Desktop-Suchmaschinen auch die Suche in speziellen Anwendungen wie E-Mail-Client und Instant Messenger unterstützen. Für den Nutzer entsteht so Mehrwert, dass auch der Inhalt von E-Mails und Protokollen vergangener Chat-Sitzungen via Instant Messenger auffindbar wird. Zudem wird getestet, ob die Programme über eine Funktion verfügen, die dem Benutzer neben der Desktop-Suche auch eine Suche im Internet ermöglichen.

Wird eine Datei gelöscht, so befindet sie sich vor der endgültigen Entfernung von der Festplatte zunächst im Papierkorb. Im Softwaretest wird geprüft, ob diese bereits für den Benutzer als gelöscht markierten Objekte noch von der Desktop-Suchmaschine gefunden

¹³ Vgl. <http://desktop.google.com/de/features.html>. Abruf: 2010-06-22.

¹⁴ Vgl. <http://www.copernic.com/en/products/desktop-search/releaseslog.html>, Abruf: 2010-06-25

werden. Die Funktion kann sich für den Benutzer als sinnvoll herausstellen, besonders wenn eine Datei versehentlich gelöscht wurde.

Große Datenmengen können zu einem Archiv komprimiert werden, beispielsweise um diese per E-Mail zu versenden. Im Test wird geprüft, ob die Desktop-Suchmaschinen auch die einzelnen in einer solchen Archivdatei befindlichen Objekte finden.

3.3.3 Sicherheit

Im Zusammenhang mit großen Suchmaschinenanbietern wie Google kam es in der Vergangenheit immer wieder zu Diskussionen über mögliche Gefahren für die Privatsphäre, weil über verschiedene Google-Anwendungen Zugriff auf private Daten der Nutzer möglich sein könnte [Schü08]. Da Desktop-Suchmaschinen einen Index erstellen, der Informationen über eine Vielzahl privater Dateien eines Nutzers enthält, stellen diese Programme prinzipiell eine gute Quelle für missbräuchliches Verhalten dar.

Standpunkt des Herstellers

Da es schwer möglich ist, den Schutz der Privatsphäre konkret zu prüfen, werden die Aussagen des Anbieters auf der Homepage, in der Installationsroutine und auf evtl. vorhandenen Hilfe-Seiten im Hinblick auf o.g. Datensicherheitsaspekte untersucht.

Schutz vor unbefugten Zugriffen

Noda und Helwig erläutern die Wichtigkeit von Sicherheitsaspekten vor allem im unternehmerischen Kontext [NoHe05, 2]. Jedoch spielt diese im privaten Bereich eine ebenso wichtige Rolle, da ein PC oft nicht nur von einer Person genutzt wird. Es besteht also grundsätzlich die Möglichkeit, dass unbefugte Dritte die Suchmaschine missbräuchlich verwenden, um private Daten ausfindig zu machen. Aus diesem Grund soll geprüft werden, ob die vom Benutzer besuchten Webseiten indiziert werden.

Weiterhin wird geprüft, ob Dateien, welche unter dem Zugang des Hauptbenutzers verfasst und abgespeichert wurden, auch von einem anderen Benutzerkonto aus gesucht und geöffnet werden können.

3.3.4 Darstellung der Suchergebnisse

Als letzter wichtiger Punkt werden Aspekte der Darstellung der Suchergebnisse im Kriterienkatalog untersucht.

Nimmt man die Windows 7-Suche als Ausgangsbeispiel, so erkennt man, dass hier Dateiname, Speicherort und das letzte Veränderungsdatum bzw. Erstellungsdatum der Suchergebnisse standardmäßig angezeigt werden. Im Test soll überprüft werden, ob diese Anzeige auch für die übrigen Suchmaschinen möglich ist.

Weiterhin wird getestet, ob die Desktop-Suchmaschinen über eine Vorschaufunktion verfügen bzw. ob Ergebnisdateien direkt im Programm geöffnet werden können, ohne dass eine weitere Anwendung gestartet werden muss.

Außerdem wird untersucht, ob die Suchmaschinen das Hervorheben der Suchterme in der Ergebnismenge durch farbliche Unterlegung (Highlighting) unterstützt [BuSc06, 8].

Essenziell für die Akzeptanz der Ergebnisse einer Suchmaschine sind die verwendeten Rankingmechanismen. Folglich soll untersucht werden, ob die Kriterien, die über die Position eines Treffers in der Suchergebnisliste entscheiden, für den Benutzer transparent dargestellt werden. Bestandteil dieser Evaluation sind auch weitere angebotene Rankingmöglichkeiten, die der Benutzer, soweit vorhanden, anstelle der voreingestellten Kriterien verwenden kann.

4 Vorstellung der Desktop-Suchmaschinen

Bevor der eigentliche Vergleich erfolgt, werden die in die empirische Untersuchung einbezogenen Suchmaschinen im Folgenden kurz vorgestellt.¹⁵ Außerdem werden die Gründe für ihre Auswahl erläutert.

4.1 Windows 7-Suche

Im Oktober 2009 kam das aktuelle Microsoft-Betriebssystem Windows 7 auf den Markt, welches nach grundlegenden Überarbeitungen den Anwendern den Umgang mit dem PC in vielerlei Hinsicht erleichtern sollte.¹⁶ Das Betriebssystem ist deutlich schneller, Ressourcenschonender und wichtige Funktionen sind durch weniger Bedienungsaufwand als in den Vorgängerversionen erreichbar.¹⁷

In einem von der Zeitschrift „*Computerbild*“ durchgeführten Test im Oktober 2009 wurde die die Windows 7-Suche als Testsieger ermittelt.¹⁸ Die Suchfunktion wurde hier mit Copernic Desktop Search, Hulbee Desktop und A-Z Finder verglichen.

4.2 Copernic Desktop Search

In der als Ausgangsbasis für die Untersuchung dienende Benchmark-Studie von Noda und Helwig erzielte das Programm Copernic Desktop Search das beste Testergebnis [NoHe05, 5]. Die Desktop-Suchmaschine wurde im Jahr 2005 in der Version 1.5 Beta getestet und als besonders homogen bewertet.

Auch aktuell spielen die Programme des Softwareanbieters Copernic noch eine wichtige Rolle auf dem Markt der Desktop-Suchmaschinen. Die Professional-Version des Tools wurde beispielsweise von der Zeitschrift „*PC-Magazin*“ in der Ausgabe März 2009 als Sieger getestet [Nefz09]. Die Schnellebigkeit solcher Tests wird durch das Resultat des

¹⁵ Screenshots der Benutzeroberflächen aller vorgestellten Desktop-Suchmaschinen sind Anhang B zu entnehmen.

¹⁶ <http://www.microsoft.com/presspass/press/2009/oct09/10-22Windows7PR.mspx>, Abruf: 2010-06-08

¹⁷ Vgl. <http://www.heise.de/newsticker/meldung/Windows-7-Microsofts-Hoffnungstraeger-am-Start-835120.html>, Abruf: 2010-06-08.

¹⁸ Vgl. <http://www.computerbild.de/artikel/cb-News-Software-COMPUTER-BILD-Test-Windows-7-sucht-am-schnellsten-4787973.html>, Abruf am 2010-06-08.

von der „*PC Praxis*“ in der Ausgabe Mai 2009 vorgestellten Vergleichs deutlich, wo Copernic lediglich den letzten Platz unter insgesamt vier Produkten erreichte.¹⁹

In vorliegenden Suchmaschinen-Vergleich wird Copernic Desktop Search in der zum Zeitpunkt der Testphase aktuellen Version 3.2.3 untersucht.

4.3 Google Desktop

Google als Suchmaschinenbetreiber, der mit ca. 90% Marktanteil in Deutschland an der Spitze der Websuchmaschinen steht,²⁰ stellt neben zahlreichen Produkten im Kontext der Websuche, wie beispielsweise Google Earth und dem Browser Google Chrome, auch eine Desktop-Suchmaschine zur Verfügung.²¹ Bei der zu testenden Version handelt es sich um Google Desktop 5.9. Die Suchmaschine wurde ebenfalls in der Benchmark-Studie 2005 geprüft und erreichte hier den fünften Platz von zwölf untersuchten Programmen [NoHe05, 7]. Zwischenzeitlich erfolgte 2009 eine Beurteilung von Google Desktop Search durch die Zeitschrift „*PC-Welt*“ im Artikel „Die 10 besten Desktop-Suchtools“ [Rems09].

4.4 Hulbee Desktop

Hulbee Desktop ist das neueste unter den getesteten Produkten. Seit Juli 2009 ist die Suchmaschine zum Download verfügbar.²² Entwickler ist die Schweizer Firma Hulbee AG, welche bis zum August unter dem Namen Grossbay tätig war.²³ Markantestes Unterscheidungsmerkmal sowohl der Internet- als auch der Desktop-Suchmaschine ist die Visualisierung von semantisch zur Suchanfrage gehörigen Begriffen in einer Data-Cloud, die zur Verfeinerung der Suchanfrage genutzt werden können.²⁴ So ist mit Hulbee Desktop die Suche nach themenverwandten Dateien wesentlich einfacher, da diese während der Index-Erstellung anhand von Kontext-Informationen kategorisiert werden.

¹⁹ Vgl. <http://www.testberichte.de/a/weiteres-tool/magazin/pc-praxis-5-2009/154592.html>. 2009, Abruf am 2010-06-08.

²⁰ Vgl. <http://www.webhits.de/deutsch/index.shtml?webstats.html>, Abruf: 2010-10-28.

²¹ Vgl. <http://www.google.de/options/>, Abruf: 2010-06-08.

²² Vgl. [http://company.hulbee.com/news/?tx_ttnews\[pointer\]=1&tx_ttnews\[tt_news\]=28&tx_ttnews\[backPid\]=28&cHash=f55fb0161f](http://company.hulbee.com/news/?tx_ttnews[pointer]=1&tx_ttnews[tt_news]=28&tx_ttnews[backPid]=28&cHash=f55fb0161f), Abruf:2010-06-20.

²³ Vgl. [http://company.hulbee.com/news/?tx_ttnews\[pointer\]=1&tx_ttnews\[tt_news\]=30&tx_ttnews\[backPid\]=28&cHash=1ea66703d2](http://company.hulbee.com/news/?tx_ttnews[pointer]=1&tx_ttnews[tt_news]=30&tx_ttnews[backPid]=28&cHash=1ea66703d2), Abruf am 2010-06-20.

²⁴ http://company.hulbee.com/products_and_solutions/hulbee/,
http://company.hulbee.com/products_and_solutions/hulbee_desktop/, Abruf am 2010-06-08.

Da die Desktop-Suchmaschine relativ kurz auf dem Markt ist, wurde sie bisher nur, wie bereits die Windows 7-Suche, im Test der „*Computerbild*“ erwähnt.²⁵ Erfahrung im Bereich der Desktop-Suche kann der Anbieter dennoch vorweisen, da es sich bei Hulbee Desktop um den Nachfolger des von der Grossbay AG angebotenen Desktop-Suchtools Superior Search handelt.²⁶ Getestet wird Hulbee in der aktuellen Version 1.0.2.12.

4.5 Xfriend personal Desktop Search

Die Desktop-Suchmaschine xfriend wird von der xdot GmbH vertrieben, einer Tochtergesellschaft des Frankfurter Software- und Beratungshauses Convotis AG.²⁷ Das getestete Produkt xfriend personal wird in der aktuellen Version 2.9 vom Anbieters als 30-Tage Testversion zur Verfügung gestellt und kann als Download-Version für 19,95 € oder als CD-Version für 23,95 € im Online-Shop erworben werden.²⁸ Auf der eigenen Internetseite werden insgesamt sechs Auszeichnungen, u.a. von *PC Welt* und *Heise Online*, mit den Urteilen „sehr gut“ bzw. „gut - sehr gut“ für die Suchmaschine präsentiert.²⁹ Im bereits erwähnten Test der „*PC Praxis*“ erzielte xfriend das beste Ergebnis.³⁰

4.6 Archivarius 3000

Mit Archivarius 3000 von Likasoft wird ein weiteres Programm in den Suchmaschinen-Test übernommen, welches bereits Teil der Studie von Noda und Helwig war [NoHe05, 5 f.]. Die Desktop-Suchmaschine belegte hier den dritten Platz hinter Copernic Desktop Search und dem damals noch verfügbaren Programm Yahoo! Desktop Search. Anstelle der Version 3.14, die in der Benchmark-Studie als Testgegenstand diente, wurde die vom Anbieter zur Verfügung gestellte 30-Tage Testversion in der Version 4.30 getestet.³¹ Die

²⁵ Vgl. <http://www.computerbild.de/artikel/cb-News-Software-COMPUTER-BILD-Test-Windows-7-sucht-am-schnellsten-4787973.html>, Abruf am 2010-06-08

²⁶ Vgl. [http://company.hulbee.com/news/?tx_ttnews\[pointer\]=1&tx_ttnews\[tt_news\]=28&tx_ttnews\[backPid\]=28&cHash=f55fb0161f](http://company.hulbee.com/news/?tx_ttnews[pointer]=1&tx_ttnews[tt_news]=28&tx_ttnews[backPid]=28&cHash=f55fb0161f), Abruf:2010-06-20.

²⁷ Vgl. <http://www.xfriend.de/de/unternehmen/ueber-xdot-gmbh/>, Abruf: 2010-06-08.

²⁸ Vgl. <http://www.xfriend.de/de/privat/uebersicht/>, <https://www.xfriend.de/de/shop/start/>, Abruf am 2010-06-08.

²⁹ <http://www.xfriend.de>, Abruf: 2010-06-08.

³⁰ Vgl. <http://www.testberichte.de/a/weiteres-tool/magazin/pc-praxis-5-2009/154592.html>. 2009, Abruf: 2010-06-08.

³¹ Vgl. <http://www.likasoft.com/de/download.shtml>, Abruf: 2010-06-08.

Vollversion mit Lizenz kann direkt über die Likasoft-Internetseite zum Preis von 29,95 € (Studenten 19,95 €) erworben werden.³²

4.7 Find and run robot (FARR)

Die Software Find and run robot unterscheidet sich in einigen wesentlichen Punkten von den anderen getesteten Programmen.³³ Es handelt sich zwar in erster Linie ebenfalls um eine Desktop-Suchmaschine, aber es werden eine Reihe zusätzlicher Optionen angeboten. So kann Find and run robot³⁴ mittels Befehlseingabe beispielsweise als Programmstarter verwendet werden. Verschiedene Internet-Suchen können direkt über das Eingabefeld des Tools durchgeführt werden. Ebenfalls ungewöhnlich ist die Bedienung ausschließlich über die Tastatur. Als einziges der getesteten Programme legt FARR keinen Index an, auf den bei der Anfrageverarbeitung zurückgegriffen werden kann [Depa09].

Interessenten können sich das Tool auf der Internetseite www.donationcoder.com herunterladen.³⁵ FARR ist entgeltfrei, es handelt sich jedoch um sogenannte Donationware, das heißt, Nutzer der auf der Seite angebotenen Software können eine freiwillige Spende in beliebiger Höhe geben und erhalten so verschiedene exklusive Vorteile. Getestet wird die Version 2.86.01.

³² Vgl. <http://www.likasoft.com/de/order.shtml>, Abruf: 2010-06-08.

³³ Vgl. <http://www.donationcoder.com/Software/Mouser/findrun/index.html>, Abruf am 2010-06-08.

³⁴ Im Folgenden wird anstelle des kompletten Programmtitels häufiger die Abkürzung FARR verwendet.

³⁵ <http://www.donationcoder.com/index.html>, Abruf: 2010-06-08.

5 Testumgebung

Voraussetzung für einen objektiven Suchmaschinen-Vergleich ist das Bereitstellen einer einheitlichen Testumgebung, die über die gesamte Evaluation hinweg verwendet wird. Nur so können gleiche Ausgangsbedingungen und damit vergleichbare Resultate gewährleistet werden.

5.1 Technische Daten

Da die Desktop-Suchmaschinen in ihrer jeweils aktuellsten Version getestet werden, wird auch bei der IT-Struktur der Testumgebung auf Aktualität geachtet, so dass keine verfälschten Ergebnisse durch die Verwendung von veralteter Technik entstehen können. Darüber hinaus kann von einer gewissen Computer- und Technikaffinität eines potentiellen Desktop-Suchmaschinen-Nutzers ausgegangen werden und damit die Verwendung aktueller Hard- und Software begründet werden. Als Betriebssystem wird Microsoft Windows installiert, da es mit über 90% Marktanteil³⁶ am häufigsten verwendet wird und somit am wahrscheinlichsten das System eines Standardnutzers repräsentiert. Auch hier wird mit Windows 7 die aktuellste Version verwendet. In der nachfolgenden Tabelle sind die Komponenten des Testsystems zusammengefasst:

Komponente	Bezeichnung
PC	Asus Barebone AM2+, Terminator3-M3N8200
CPU	AMD Athlon64 X2 5000+
RAM	GeIL DIMM 4 GB DDR2-1066 Kit
Festplatte	Samsung HD252HJ 250 GB, 3,5 Zoll SATA,
Betriebssystem	Microsoft Windows 7 Professional, 64 Bit-Version

Tab. 1: Komponenten des Testsystems

Nach Abschluss des Testvorgangs für eine Suchmaschine findet stets eine Neuinstallation des Betriebssystems statt. Um realistische Ausgangsbedingungen für einen Benutzer-PC zu schaffen, werden die in Tabelle 2 zusammengefassten Anwendungsprogramme jeweils standardmäßig installiert.

³⁶ Vgl. <http://marketshare.hitslink.com/operating-system-market-share.aspx?qprid=9>, Abruf: 2010-06-20.

Softwareart	Programmname	Version
Flash-Player	Adobe Flash Player	10.0.45.2
Bildbearbeitung	Adobe Photoshop	7.0
Antivirenprogramm	Avast! Free Antivirus	5.0.545.0
Instant Messenger	ICQ	7.1
Browser	Mozilla Firefox	3.6.3
E-Mail-Client	Mozilla Thunderbird	3.0.4
Textverarbeitung u.ä.	Open Office	3.2.9483
VoIP-Software	Skype	4.2.155

Tab. 2: Anwendungssoftwarekollektion des Testsystems

5.2 Testkollektion

Als Hauptaspekt für die Komposition der Testkollektion steht die Schaffung einer möglichst realistischen Testumgebung für die durchzuführenden Tests im Vordergrund. Als Ausgangsbasis für diese umfangreiche Testdatenbasis dient die digitale Bibliothek des Fachgebiets Informations- und Wissensmanagement der TU Ilmenau, welche überwiegend wissenschaftliche Texten im PDF-Format enthält, wodurch sich der verhältnismäßig hohe Anteil der PDF-Dokumente in Tabelle 3 erklärt, in welcher der erste Teil der in der Testkollektion enthaltenen Dateien zusammengefasst ist.

Da Mediadateien einen großen Teil der auf Benutzer-PCs gespeicherten Datenmenge repräsentieren, ist ebenfalls ein Anspruch dieser Untersuchung, das Verhalten der Desktop-Suchmaschinen im Umgang mit Audio- und Videodateien zu analysieren. Die verschiedenen Formate, die die Testkollektion vervollständigen, sind in Tabelle 4 zusammengestellt.

Darüber hinaus befinden sich im Posteingang von Mozilla Thunderbird zu jedem Zeitpunkt des Softwaretests ca. 1000 E-Mails. Diese werden indexiert, soweit die getestete Suchmaschine den E-Mail-Client unterstützt.³⁷

³⁷ Ergänzende Bemerkungen zu Testsystem und -kollektion sind Anhang A zu entnehmen.

Datenformat	Anzahl	Datenformat	Anzahl	Datenformat	Anzahl
.pdf	1970	Videodateien		Quellcodedateien u.a.	
.doc	111	.mpg	30	.htm	95
.ppt	94	.avi	27	.css	94
.xls	19	.mp4	27	.html	48
.txt	16	.wmv	25	.mht	35
.psd	5	.mkv	22	.h	17
.docx	1	.swf	22	.c	10
.pps	1	.mov	21	.js	10
Grafikdateien		.flv	17	.myd	4
.gif	110	Audiodateien		.eps	2
.jpg	83	.mp3	89	.chm	1
.bmp	37	.aac	25	.ps	1
.ico	6	.ogg	25	.opt	1
Archivdateien u.a.		.m4a	23		
.zip	26	.au	18		
.rar	4	.aiff	17		
.tar	2	.wav	14		
.iso	2	.w64	13		
Ausführungsdateien					
.exe	10				

Tab. 3: Dateiformate der Testkollektion

6 Suchmaschinen-Vergleich

In diesem Kapitel erfolgt die Darstellung der Ergebnisse des Software-Tests, welcher mit Hilfe der Kapitel drei vorgestellten Kriterienkatalog durchgeführt wurde. Die Kriterien werden nachfolgend um genauere Informationen hinsichtlich der methodischen Aspekte bzgl. Messungsdurchführungen, Auswahl der Suchanfragen, u.ä. ergänzt.

6.1 Vielseitigkeit & Einschränkungen

In Tabelle 5 werden die untersuchten Einzelkriterien zur Ermittlung der Vielseitigkeit und eventueller Einschränkungen der Desktop-Suchmaschinen dargestellt. Die Spaltenbezeichnung entspricht dem Namen des jeweils getesteten Programms. Mittels der verwendeten Symbolik kommt eine positive, negative oder neutrale Bewertung zum Ausdruck.³⁸

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
Deutschsprachige Version	+	+	+	+	+	+	-
Versionen für verschiedene Betriebssysteme	-	+	-	+/-	-	-	-
Add-Ons	+	-	-	-	-	-	+
Einschränkungen der Testversion	+	+	+	+	-	+	+
Σ	3	3	2	2,5	1	2	2

Tab. 4: Vielseitigkeit und Einschränkungen

Sechs der sieben Desktop-Suchmaschinen sind in einer deutschsprachigen Version verfügbar. Lediglich Find and run robot ist nur in der englischen Version erhältlich. Die Windows 7-Suchfunktion verwendet automatisch die Sprache, die als Betriebssystemsprache eingestellt ist, also in unserem Fall für das bei der Untersuchung verwendete Testsystem Deutsch.

Eine Windows-kompatible Version der Desktop-Suchmaschine wird selbsterklärend für die Windows 7-Suche angeboten und darüber hinaus auch für alle anderen getesteten Programme. Für Google Desktop, Copernic Desktop Search, xfriend, sowie FARR wird

³⁸ Diese Tabellenform, wie für Tab. 5 verwendet, wird innerhalb des 6. Kapitels als Standardform für eine Reihe von Auswertungen verwendet. Die genaue Interpretation der angewandten Symbolik ist Anhang C zu entnehmen. Im vorliegenden Abschnitt wird häufig auf Angaben auf den Webseiten der Anbieter Bezug genommen. Die Webadressen können ebenfalls Anhang C entnommen werden.

ausdrücklich die Kompatibilität mit Windows 7 gewährleistet, während bei Hulbee und Archivarius3000 für die getesteten Versionen noch kein Verweis auf die Verwendbarkeit mit dem neuen Betriebssystem zu finden ist.

Google Desktop ist als einzige der Suchmaschinen in drei Betriebssystem-Versionen für Windows, Mac und Linux verfügbar. Für xfriend steht neben der Windows- auch eine Linux-Version bereit. Alle anderen Programme sind nur für eine Verwendung mit Windows geeignet.

Kostenlose Zusatzfunktionen sind für die Windows 7-Suche und FARR verfügbar. Das Tool Start ++ ergänzt das Suchfeld im Windows-Startmenü um hilfreiche Funktionen.³⁹ Ähnlich wie bei FARR kann über das Sucheingabefeld eine Suche in Google oder Wikipedia durchgeführt werden. Auch das Öffnen von Programmen mittels Befehlseingabe ist über das vom Windows-Entwickler Brandon Paddock entworfene Tool möglich. Eine Version, die mit Windows 7 kompatibel sein soll, ist bereits verfügbar [Padd09]. Für FARR werden auf www.donationcoder.com mehr als 50 Plug-In angeboten, über die die Suchmaschine um zusätzliche, den persönlichen Vorlieben entsprechenden, Funktionen erweitert werden kann.⁴⁰ Beispielsweise wird ein Add-On angeboten, mit dessen Hilfe bei einer Suchanfrage auch die Favoritenliste des Firefox-Browsers durchsucht wird.

Da alle Produkte bis auf xfriend und Archivarius 3000 in der kostenlosen Vollversion getestet werden, beschränkt sich die Frage der Einschränkungen auf diese beiden Suchmaschinen. Für beide lässt sich zunächst festhalten, dass die Testversion eine Gültigkeit von 30 Tagen aufweist. Die Testversion von Archivarius ist außerdem gegenüber der Vollversion eingeschränkt nutzbar, da nur 10000 Dateien indexiert werden können. Sobald dieses Maximum an Dateien verzeichnet wurde, wird der Vorgang automatisch abgebrochen und der Benutzer erhält über ein Pop-up-Fenster den diesbezüglichen Hinweis. Aus diesem Verhalten der Suchmaschine entstehen natürlich Beeinträchtigungen für weitere Untersuchungsgegenstände, da beispielsweise die Indexgröße oder auch Retrievalmaße im Vergleich zu den anderen Suchmaschinen, die alle Testdaten komplett indexieren, nicht mehr sinnvoll vergleichbar ist. An den entsprechenden Stellen des Software-Tests wird hierauf entsprechend eingegangen. Auf der Homepage von Likasoft ist kein Hinweis zu finden, der

³⁹ Vgl. <http://www.drwindows.de/windows-news/188-start-maechtige-erweiterung-fuer-das-suchfeld-startmenue.html>, Abruf: 2010-06-22.

⁴⁰ Vgl. <http://www.donationcoder.com/Software/Mouser/findrun/addons/>, Abruf: 2010-06-22

vor dem Download und der Installation des Programms über diesen Sachverhalt informiert!

Für die Suchmaschine xfriend werden ebenfalls keine Einschränkungen der Testversion auf der Internetseite angegeben, allerdings kann während des Tests auch nichts in dieser Hinsicht festgestellt werden. Allein gelegentlich eingeblendete Werbebanner mit dem Slogan „Kaufen Sie jetzt die Vollversion!“, z.B. oberhalb der Suchergebnisse, machen darauf aufmerksam, dass man mit einer zeitlich beschränkten Testversion arbeitet.

6.2 Effizienz

Nachfolgend werden die Belastung von Festplattenspeicher, Hauptprozessor und Arbeitsspeicher durch die einzelnen Programme getestet.

Größe von Installationsdatei und Programmordner

In der nachfolgenden Grafik werden die Größe der herunterzuladenden Installationsdatei, sowie die Größe des nach der Installation angelegten Programmordners, für jede Suchmaschine dargestellt.⁴¹

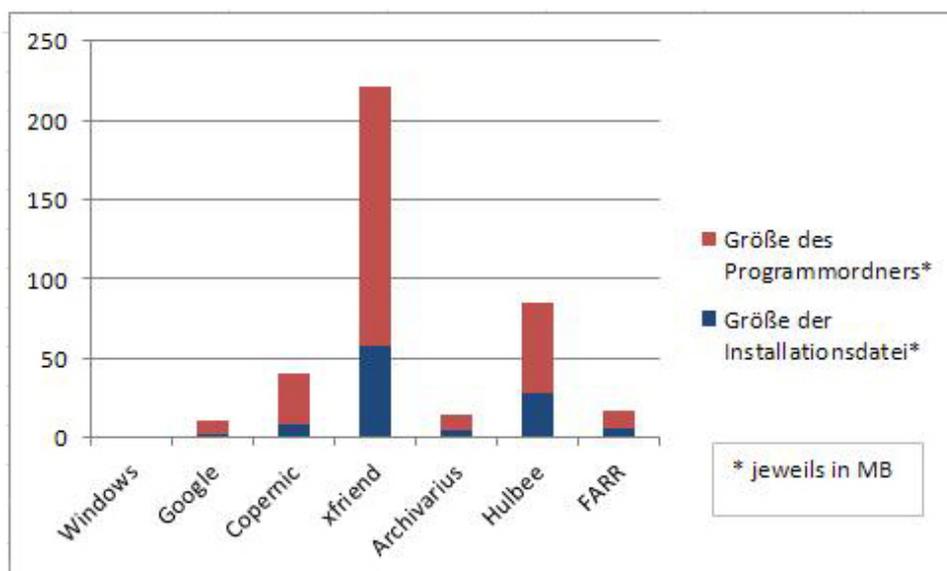


Abb. 2: Größe von Installationsdatei und Programmordner

⁴¹ Eine Tabelle mit den genauen Messwerten, sowie Details zur Grundlage der Messung, sind Anhang D zu entnehmen.

Betrachtet man die Größe der Installationsdatei, so lassen sich große Unterschiede zwischen den einzelnen Desktop-Suchmaschinen feststellen. Auf der einen Seite stehen die Programme mit einer relativ geringen Downloadgröße von unter 10 MB, wozu Google Desktop, Archivarius 3000, FARR und Copernic gehören. Im Vergleich benötigen Hulbee Desktop mit 27,9MB und xfriend mit 58,0MB relativ viel Speicher. Für die Windows-Suche kann in diesem Zusammenhang keine sinnvolle Messung durchgeführt werden, da sie nicht heruntergeladen werden muss.

Auch für die Größe des nach der Installation angelegten Programmordners kann für die Windows 7-Suche keine Angabe gemacht werden. Die Suchfunktion ist ein sogenannter Windows Dienst und unter dem Pfad C:/Windows/System32/Dienste/Windows Search zu finden. Für die anderen Suchmaschinen bleibt die eben beschriebene Rangfolge bestehen, da die Größen von Installationsdatei und Programmverzeichnis parallel ansteigen. Das heißt, je größer die Installationsdatei, desto größer ist auch das installierte Programmverzeichnis.

Speicherauslastung

Abbildung 3 und 4 illustrieren die Auslastung von CPU und Arbeitsspeicher⁴² während des

- alleinigen Programmbetriebs (P), während
- einer Suchanfrage (S) und während
- der Indexierung (I).⁴³

⁴² In den Abbildungen werden die Abkürzungen „CPU“ für den Hauptprozessor und „RAM“ für den Arbeitsspeicher verwendet.

⁴³ Eine Tabelle mit den genauen Messwerten, sowie Details zur Durchführung der Messung, sind Anhang D zu entnehmen.

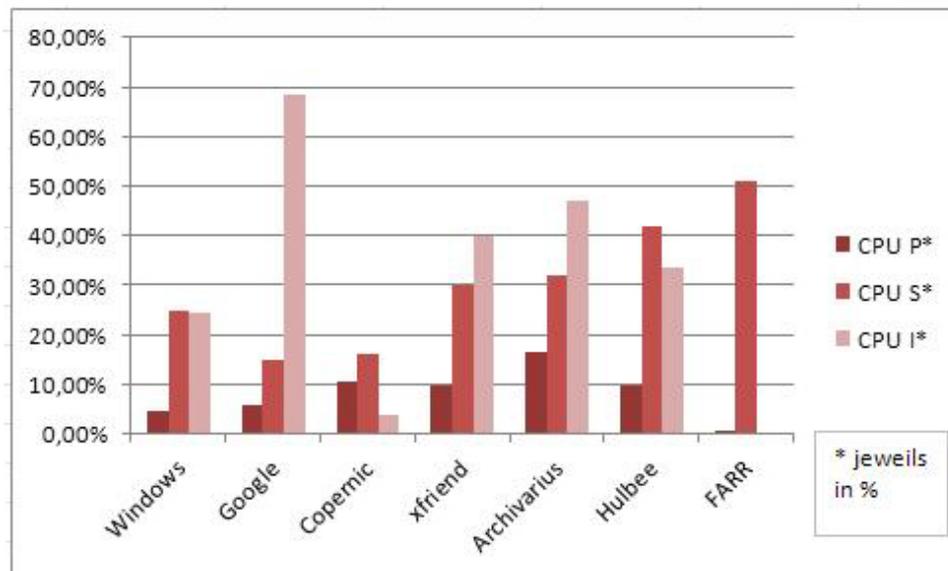


Abb. 3: Auslastung der CPU

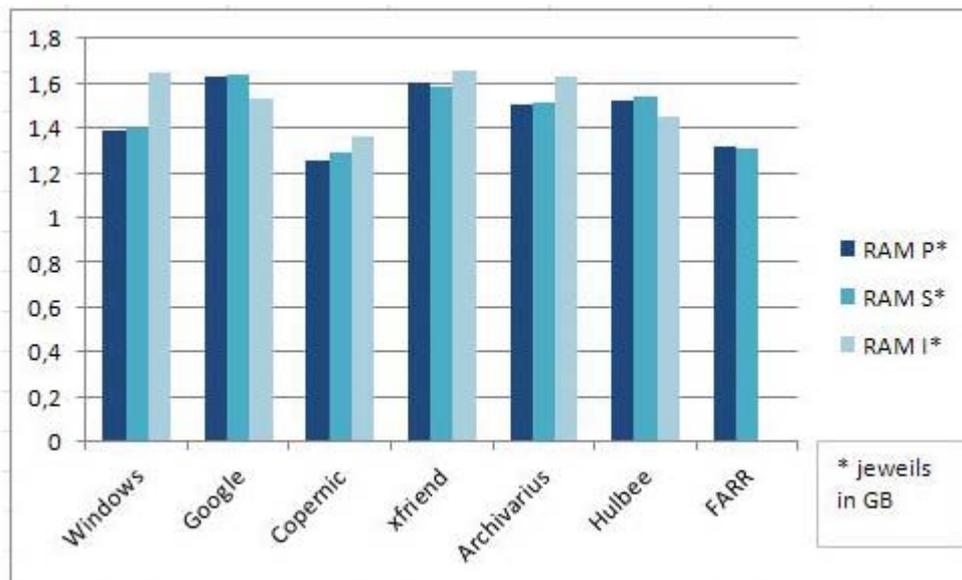


Abb. 4: Auslastung des Arbeitsspeichers

Programmbetrieb

Unter immer gleichen Ausgangsbedingungen gelangt man zum Ergebnis, dass CPU- und Arbeitsspeicherauslastung nicht linear verlaufen. Google Desktop Search weist beispielsweise den höchsten Messwert bzgl. der durchschnittlichen RAM-Auslastung auf. Betrachtet man jedoch die CPU-Auslastung, so liegt Google hier im Mittelfeld und es finden sich vier Suchmaschinen mit höheren Auslastungswerten.

Während des Tests ist Archivarius 3000 besonders aufgefallen, da bei dieser Suchmaschine die CPU-Auslastung starken Schwankungen unterworfen war und sie stetig nach wenigen

Sekunden wechselte, i.d.R. von einem sehr niedrigen zu einem sehr hohen Wert und umgekehrt. Bei allen anderen Suchmaschinen veränderte sich die Prozessor-Auslastung in einer weitaus geringeren Frequenz, in etwa einmal pro Minute. Da folglich für die restlichen Suchmaschinen auch die Anzahl der über den Beobachtungszeitraum gemessenen Werte geringer ausfällt, weist Archivarius den höchsten CPU-Auslastungswert auf. Den geringsten CPU-Bedarf hat FARR mit lediglich 0,5%.

Allgemein betrachtet fallen die Schwankungen der Werte zwischen den einzelnen Suchmaschinen im Programmbetrieb am geringsten aus, wenn man die Messwerte mit denen des Suchvorgangs oder der Indexierung vergleicht. Besonders deutlich wird es, am Beispiel der CPU-Auslastung.

Suchvorgang

Hinsichtlich der Arbeitsspeicherauslastung treten im Vergleich mit dem im letzten Abschnitt erläuterten Betriebszustand der Suchmaschine keine nennenswerten Veränderungen auf. Die gemessenen Werte weichen nur in einem minimalen Spektrum von 0,01 bis 0,04 GB nach oben ab. Die Rangfolge der Suchmaschinen bleibt hierdurch unverändert.

Anders hingegen verhält es sich mit der CPU-Auslastung nach Start der Suchanfrage „Informationsmanagement“. Der geringste Auslastungswert kann mit 15% bei Verwendung von Google Desktop Search gemessen werden. Danach folgen Copernic, Windows, xfriend, Archivarius3000 und Hulbee.

Eine Ausnahme vom allgemeinen Vorgehen bei der Messung muss für FARR gemacht werden, da diese Suchmaschine eine Live-Suche durchführt.⁴⁴ Hierbei wird bereits während der Eingabe der Anfrage mit der Suche begonnen, was die Bestätigung mit „Enter“ unnötig macht. Aus diesem Grund wird ersatzweise der höchste auftretende Wert während der Suchbegriff-Eingabe gemessen. Dieser beläuft sich auf 51%.

Indexierung

Betrachtet man die RAM-Auslastung während des Indexierungsvorgangs, so erhält man eine andere Rangfolge der Suchmaschinen, als für die letzten beiden getesteten Kennzahlen. Den besten Wert erreicht Copernic mit einer durchschnittlichen Auslastung von 1,36 GB ab. Dieser Wert kann für beide von Copernic angebotenen Indexierungsgeschwindig-

⁴⁴ Vgl. <http://www.donationcoder.com/Software/Mouser/findrun/index.html>, Abruf: 2010-06-08.

keiten gemessen werden. Der Benutzer hat bei dieser Suchmaschine die Möglichkeit, die Performance der Indexierung manuell einzustellen, wobei zwischen einer unbegrenzten Variante, die alle Computerressourcen nutzt, und einer eingeschränkten Variante, die für geringen Ressourcenbedarf optimiert ist, gewählt werden kann. Für Find and run robot kann keine Angabe gemacht werden, da die Suchmaschine, wie schon in Kapitel 4 beschrieben, keinen Index erstellt, sondern bei einer Suchanfrage die einzelnen Verzeichnisse durchsucht.

Wie aus Abbildung 3 hervorgeht, können die größten Schwankungen zwischen den einzelnen Suchmaschinen insgesamt für die CPU-Auslastung während der Indexierung betrachtet festgehalten werden. Sehr auffällig ist der hohe Auslastungswert von Google Desktop-Search, welche im Gegenzug für Programmbetrieb und Suchvorgang eine sehr niedrige Auslastung verzeichnen kann.

6.3 Usability

Die Windows 7-Suche erlaubt dem Benutzer, durch die Verankerung der Sucheingabefelder im Startmenü und im Windows-Explorer sehr einfach und jederzeit auf die Eingabefunktion zugreifen. Die Darstellungsweise wirkt insgesamt sehr übersichtlich. Nachteilig ist, dass einige Funktionen im Vergleich mit anderen Suchmaschinen sehr versteckt liegen. Die Rankingmöglichkeiten werden beispielsweise nur durch einen Rechtsklick in den freien Raum und anschließende Auswahl von „Sortieren nach“ sichtbar. Allerdings ist diese Funktion auch aus dem ‚normalen‘ Arbeiten mit dem Windows-Explorers bekannt.

Sehr einfach zu bedienen ist Google Desktop Search für Benutzer, die Google bereits als Web-Suchmaschine verwenden, denn der Aufbau von Desktop- und Internet-Suche ist nahezu identisch. Durch die Vielzahl an angebotenen Darstellungsformen (s. Kapitel 6.10) ist ebenso einfach auf die Desktop-Suche zuzugreifen. Eine Besonderheit, die Google gegenüber den anderen getesteten Programmen auszeichnet, ist die zusammen mit der Desktop-Suche installierte Seitenleiste, auf welcher sich u.a. eine Uhr, aktuelle News und eine Wettervorhersage befinden. Der Nutzer kann aus einer Vielzahl sogenannter Google Gadgets weitere Applikationen auswählen, die den persönlichen Präferenzen entsprechen, und diese der Seitenleiste hinzufügen.⁴⁵

⁴⁵ Vgl. <http://desktop.google.com/de/features.html>, Abruf: 2010-06-22.

Bei Copernic fällt zunächst positiv auf, dass die Ergebnisse durch eine Unterteilung für den Benutzer sehr übersichtlich und strukturiert dargestellt werden. Für jede dieser Kategorien wie E-Mail, Dateien oder Musik gibt es eigene spezifische Suchoptionen. So kann beispielsweise in der Kategorie E-Mail nicht nur nach dem Betreff und Inhalt der Poststücke, sondern auch nach deren Absender gesucht werden.

Xfriend bietet, im Vergleich zu Google Desktop oder der Windows 7-Suche, viele Funktionen auf der Oberfläche an und wirkt dennoch sehr strukturiert. Positiv ist die Nummerierung der Suchergebnisse anzumerken, die xfriend als einzige Suchmaschine neben FARR verwendet, sowie die Anzeige von Weiterverarbeitungsmöglichkeiten der Suchergebnisse, z.B. Ausdrucken oder Speichern, über kleine Symbole direkt neben dem jeweiligen Ergebnis.

Öffnet man die Suchmaschine Archivarius fällt zunächst das sehr schlichte Design auf. So wurden beispielsweise keine Farben verwendet, was v.a. im Vergleich mit der von gelb und orange dominierten Copernic Desktop Search oder Hulbee auffällt. Dieser Sachverhalt kann durchaus die Orientierung des Benutzers im Programm beeinflussen. So fällt das Sucheingabefeld bei letztgenannten Suchmaschinen sofort ins Auge, bei Archivarius hingegen nicht. Positiv zu bemerken ist die Indexverwaltung, die im Top-Menü mit der Suchfunktion hierarchisch gleichgestellt ist. Bei keiner anderen der getesteten Desktop-Suchmaschinen ist der Zugriff auf Index und Indexierungsoptionen so unmittelbar möglich.

Die Benutzeroberfläche von Hulbee ist ähnlich minimalistisch eingerichtet wie Google Desktop Search. Genau wie für xfriend existiert auch für Hulbee ein Benutzerhandbuch als PDF-Dokument. Einen großen Vorteil für den Benutzer und Alleinstellungsmerkmal unter den getesteten Suchmaschinen ist die bereits im vierten Kapitel erläuterte Visualisierung von Suchergebnissen mittels Data Cloud.

Find and run robot ist die auf den ersten Blick minimalistischste Desktop-Suchmaschine, denn sie besteht nur aus einem Sucheingabefeld bzw. einer Kommandozeile. Der Bereich zur Darstellung der Suchergebnisse (vgl. Screenshot der FARR-Benutzeroberfläche in Anhang B), wird erst nach einer Suchanfrage angezeigt. Eine Menüleiste fehlt komplett. Lediglich über einen kleinen Button neben dem Suchfeld gelangt man zu den Einstellungen. Die so zu öffnende Optionsauswahl ist allerdings sehr umfangreich, und, weil in Englisch, nicht für jeden Nutzer unmittelbar verständlich. Eine Hilfefunktion zur Orientie-

rung ist verfügbar. FARR ist die einzige getestete Desktop-Suchmaschine, die dem Benutzer die Möglichkeit bietet, selbstständig Veränderungen am Design vorzunehmen. Im Einstellungsbereich können Hintergrundfarben und Schriftarten verändert werden (siehe ebenfalls Anhang B). Darüber hinaus kann der Benutzer beispielsweise auch die für die Bedienung über das Keyboard hilfreichen Tastenkürzel festlegen.

6.4 Sprachliche Analyse

Stemming

In Tabelle 6 werden die an die Suchmaschinen gestellten Suchanfragen in den Zeilen und die möglichen Suchergebnisse in den Spalten erfasst. Zeilen und Spalten sind identisch, da das Ziel der Untersuchung ist, herauszufinden, welche der ebenfalls „Spiel“ beinhaltenden Begriffe bei der Anfrage „Spiel“ gefunden werden, usw. Mit jeder Desktop-Suchmaschine werden alle sechs Suchanfragen durchgeführt. Sobald für die jeweilige Anfrage eine der 6 Wortvarianten als Suchergebnis angezeigt wird, wird die getestete Suchmaschine am jeweiligen Kreuzungspunkt in der Tabelle vermerkt. Zur Vereinfachung werden statt dem kompletten Programmnamen nur Abkürzungen⁴⁶ angegeben. Unabhängig von der Gesamttrefferzahl werden alle Suchergebnisse zur Ergebnisfindung berücksichtigt.

Betrachtet man die Diagonale der Tabelle von links oben nach rechts unten, so kann man feststellen, dass von jeder Suchmaschine der angefragte Suchbegriff als Suchergebnis angezeigt wird. Bemerkenswert sind jedoch die Unterschiede zwischen den Programmen, wenn es um die weiteren Suchergebnisse geht.

Die Suchmaschine mit den meisten exakten und ähnlichen Suchergebnissen ist FARR mit insgesamt 15 Nennungen.

⁴⁶ A: Archivarius, C: Copernic, F: Find and run robot, G: Google, H: Hulbee, W: Windows, X: xfriend.

		Ergebnisse					
Suchanfragen		Spiel	Fußball- spiel	Spieler	ge- spielt	Spiele- risch	spielen
	Spiel	W, G, C, X, H, A, F	F	X ,C, F	C, F	C, F	X, H, C, F
	Fußball- spiel		W, G, C, X, H, A, F				
	Spieler	X, H, A		W, G, C, X, H, A, F		X ,C, F	X, H, F
	gespielt				W, G, C, X, H, A, F		F
	Spiele- risch			H		W, G, C, X, H, A, F	
	spielen	X, H		X		F	W, G, C, X, H, A, F

Tab. 5: Stemming

Unterscheidung von Singular und Plural

Um herauszufinden, in wieweit die einzelnen Desktop-Suchmaschinen eine Unterscheidung zwischen Singular- und Pluralformen durchführen, werden acht verschiedene Suchanfragen durchgeführt, welche in den Zeilen der Tabelle 6 genannt werden.⁴⁷ Die positive Bewertung mit „+“, erfolgt in diesem Fall, wenn in den ersten zehn Suchergebnissen nur die angefragte Form oder auch zum selben Wortstamm gehörige Begriffe angezeigt werden. Lautet die Suchanfrage beispielsweise „Haus“, so wird das „+“ auch vergeben, wenn neben der exakten Form „Hausmeister“ o.ä. genannt wird. Wichtig ist, dass nicht die Pluralform „Häuser“ und Abwandlungen hiervon, wie z.B. „Hochhäuser“ in den ersten zehn Suchergebnissen genannt werden. Ist dies der Fall, so wird die Suchmaschine negativ bewertet.⁴⁸

⁴⁷ Es handelt sich um Singular- und Pluralformen der deutschen Sprache mit Ausnahme des Begriffs „Computer“ und der zugehörigen Pluralform aus dem Englischen („Computers“).

⁴⁸ Anmerkung: Bei der Suchanfrage „Kind“ darf folglich auch nicht „Kindergarten“ in den Suchergebnissen stehen. Zwar handelt es sich um einen Begriff des gleichen Wortstamms, verweist jedoch eindeutig auf die Pluralform „Kinder“. Irrelevante Treffer, die überhaupt keinen Bezug zur Suchanfrage aufweisen werden ignoriert.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
Kind	-	+	-	+	-	+	-
Kinder	+	+	+	+	-	-	+
Computer	+	+	-	+	-	+	+
Computers	+	+	+	-	-	-	-
Haus	-	+	+	+	+	+	+
Häuser	+	+	+	+	+	-	+
Apfel	-	+	-	+	+	+	+
Äpfel	-	+	-	+	+	-	+
Σ	4	8	4	7	4	4	6

Tab. 6: Unterscheidung von Singular und Plural

Auffällig ist, dass die Desktop-Suchmaschinen häufig für die Singular- und Pluralanfrage exakt die gleiche Suchergebnisliste präsentieren, wenn keine Unterscheidung vorgenommen wird. Copernic liefert sowohl für „Apfel“ als auch für „Äpfel“ 60 Treffer in der gleichen Rangfolge. Jedoch besteht insofern ein Unterschied, dass, je nach Anfrage, der passende Begriff in den Suchergebnissen durch Highlighting farblich hervorgehoben wird und der andere nicht.

Generell lässt sich sagen, dass fast alle Suchmaschinen zwischen dem am meisten differenzierten Anfragenpaar von „Haus“ und „Häuser“ unterscheiden. Hier wird für die Pluralform eine Endung angefügt und im Wortstamm findet eine Verschiebung von „a“ zum Umlaut „ä“ statt. Bei den anderen Anfragenpaaren wird die Pluralbildung nur durch jeweils eine dieser Veränderungen vollzogen.

Google Desktop Search unterscheidet als einzige der getesteten Suchmaschinen die Singular- und Pluralformen aller vier angefragten Begriffe.

Groß- und Kleinschreibung

Für die Untersuchung der Beachtung von Groß- und Kleinschreibung werden mit jeder Suchmaschine vier Suchanfragen durchgeführt,⁴⁹ wobei als Testkriterium gilt, dass innerhalb der ersten zehn Suchergebnisse nur die angefragte Form gelistet sein darf.

Als Ergebnis kann festgehalten werden, dass keines der Programme diese Anforderung erfüllt und stets Treffer in der jeweiligen anderen Schreibweise angezeigt werden, also bei der Suche nach „apfel“ erscheinen auch Ergebnisse, die nur die großgeschriebene Variante

⁴⁹ Suchanfragen: „Apfel“, „apfel“, „IT“, „it“.

„Apfel“ beinhalten und umgekehrt. Das geschieht unabhängig davon, ob die Bedeutung des Begriffs durch die Schreibweise verändert werden kann („IT“ → „it“) oder gleich bleibt („Apfel“ → „apfel“).

Besonders auffällig ist das Verhalten von Hulbee auf die Suchanfrage „it“. Man erhält null Ergebnisse, da die Suchmaschine das Wort offensichtlich als Stoppwort betrachtet und keine Suche nach solchen zulässt. Versucht man die Suchanfrage in „IT“ zu ändern, so ist das nicht möglich, da automatisch das bereits gespeicherte „it“ geschrieben wird.

Soundex

Zur Prüfung des Soundex-Algorithmus werden sechs unterschiedliche Schreibweisen des deutschen Namens „Meier“ als Suchanfrage an die Desktop-Suchmaschinen gestellt. Testkriterium ist, dass bei jeder Anfrage möglichst viele der alternativen Schreibweisen gefunden werden.⁵⁰

		Ergebnisse					
		Meier	Maier	Meyer	Mayer	Meyr	Mayr
Suchanfragen	Meier	W, G, C, X, H, A, F	F	F	C, F		F
	Maier	F	W, G, C, X, H, A, F	F		F	F
	Meyer	F	F	W, G, C, X, H, A, F		F	
	Mayer	H	H		W, G, C, X, H, A, F	H	H, F
	Meyr		F	F	F	W, G, C, X, H, A, F	
	Mayr		F		F		W, G, C, X, H, A, F

Tab. 7: Soundex

In Tabelle 7 wird ersichtlich, dass FARR mit 22 Nennungen und Hulbee mit zehn Treffern den Soundex-Algorithmus unterstützen. Neben den gesuchten Schreibweisen von „Meier“ finden diese beiden Suchmaschinen noch weitere Ableitungen des Namens, beispielsweise *Meir* und *Myers*, aber auch *Mei*, *Mai* und *May*. Alle anderen Suchmaschinen finden nur die jeweils angefragte Schreibweise (vgl. Tabellendiagonale von links oben nach rechts unten).

⁵⁰ Für die Tabelle 7 gelten die zu Aufbau und Bewertung der Tabelle 5 (Stemming) gegebenen Erläuterungen. Zur Ergebnisfindung werden wiederum alle Suchergebnisse berücksichtigt.

6.5 Qualität der Rechtschreibprüfung

In diesem Abschnitt wird untersucht, ob die Desktop-Suchmaschinen bei falscher Eingabe eines Suchbegriffs eine automatische Korrektur oder eine Rechtschreibprüfung, in welcher der Anwender auf seinen Fehler aufmerksam gemacht wird, durchführen.

Insgesamt werden für diesen Zweck vier absichtlich fehlerhafte Schreibweisen des Wortes „*Alphabet*“ als Suchanfragen gestellt [Stoc07, 306]. Das Beispiel, durch welches die bereits im Kriterienkatalog genannten häufigsten Fehlertypen repräsentiert werden, wird in Anlehnung an Stock [Stoc07, 306] zur Überprüfung der Desktop-Suchmaschinen gewählt.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
Falscher Buchstabe (<i>ALPHIBET</i>)	-	-	+	+	-	-	-
Ausgelassener Buchstabe (<i>ALPABET</i>)	-	-	+	+	-	-	+
Zusätzlicher Buchstabe (<i>ALLPHABET</i>)	-	-	+	+	-	-	+
Buchstaben-Dreher (<i>ALHPABET</i>)	-	-	+	+	-	-	-
Σ	0	0	4	4	0	0	2

Tab. 8: Rechtschreibprüfung und –korrektur

Vier der sieben Desktop-Suchmaschinen führen (in der untersuchten Version) keine Rechtschreibprüfung oder –korrektur durch. Google, die Windows-Suche, Hulbee und Archivarius visualisieren dem Benutzer nach der fehlerhaften Suchanfrage lediglich, dass keine Suchergebnisse gefunden wurden. Die vorteilhafteste Mitteilung dieser Art verwendet Archivarius mit folgender Formulierung: „Dokument nicht gefunden! Das Wort „*alphet*“ kommt nirgendwo vor.“ Durch die ausdrückliche Erwähnung der falschen Schreibweise wird der Benutzer auf seinen Fehler aufmerksam gemacht und er kann seine Eingabe entsprechend verbessern.

Copernic und xfriend führen die Rechtschreibprüfung durch. Nach Eingabe eines nicht existierenden Suchbegriffs wird von der Suchmaschine ein alternativer Vorschlag „meinen Sie: *alphabet*“ angeboten. Anders als bei xfriend kann die Option der Rechtschreibprüfung bei Copernic auch deaktiviert werden. Eine automatische Korrektur wird von FARR durchgeführt. Jedoch trifft dies, wie in Tabelle 8 zu sehen, nur auf zwei der vier Suchanfragen zu. Die Anfragen „*Alphet*“ und „*Alhpabet*“ liefern trotz falscher Schreibweise

Suchergebnisse, haben jedoch keinen Bezug zum gesuchten Begriff „*Alphabet*“. Im Gegensatz dazu werden für „*Alpabet*“ teilweise, und für „*Allphabet*“ ausschließlich, Treffer angezeigt, die die korrekte Schreibweise beinhalten.

6.6 Analyse bestimmter Suchfunktionen

Unterstützung der Operatoren AND, OR, NOT, NEAR

Um herauszufinden, ob die Desktop-Suchmaschinen die booleschen Operatoren AND, OR und NOT, sowie den Abstandsoperator NEAR unterstützen, werden vier diesbezügliche Suchanfragen gestellt.⁵¹ Wie im Abschnitt zur Unterscheidung von Singular und Plural werden irrelevante Suchergebnisse ohne erkennbaren Bezug zur Suchanfrage nicht beachtet. Eine negative Bewertung erfolgt, wenn ein Ergebnis auftritt, welches offensichtlich dem Grundgedanken des Operators widerspricht, z.B. wenn bei der Anfrage „*Apfel NOT Birne*“ ein Suchergebnis angezeigt wird, dass entweder nur den unerwünschten Begriff „*Birne*“ oder beide Begriffe enthält. Basis für die Bewertung sind alle Suchergebnisse.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
AND	+	+/-	+	+	-	+	+
OR	+	+/-	+	+	+	+	+
NOT	+	-	+	+	-	-	-
NEAR	-	-	+	+/-	-	-	+/-
Σ	3	1	4	3,5	1	2	2,5

Tab. 9: Boolesche Operatoren und Abstandsoperator

Besonders auffällig in der Suchergebnisliste von Google Desktop-Search sind die ersten beiden angezeigten Treffer, welche bei der Anfrage „*Apfel AND Birne*“ allein das Wort „*and*“ enthalten, jedoch weder „*Apfel*“ noch „*Birne*“! Die restlichen Ergebnisse enthalten beiden Suchterme. Bei Verwendung des OR-Operators fällt auf, dass nur Suchergebnisse mit beiden Begriffen gezeigt werden. Das ist kein Widerspruch, aber identisch mit den auch unter Verwendung des AND-Operators ausgegebenen Ergebnissen. Aus diesen Gründen wird für Google, wie Tabelle 9 zeigt, eine neutrale Bewertung vergeben.

Als Besonderheit von Copernic kann festgehalten werden, dass die Suchbegriffe in der Ergebnisliste farblich unterlegt werden. Auch scheint Copernic als einzige Suchmaschine

⁵¹ Suchanfragen: „*Apfel AND Birne*“, „*Apfel OR Birne*“, „*Apfel NOT Birne*“, „*Apfel NEAR Birne*“.

den NEAR-Operator zu unterstützen, da beide Begriffe des Suchterms beispielsweise im Dateinamen der Ergebnisse im Sinne der Semantik des NEAR-Operators als Kollokation vorkommen, wie beispielsweise „*Apfel_und_Birne.jpg*“.

Xfriend erhält in Zusammenhang mit dem Abstandsoperator eine neutrale Bewertung, da nur Suchergebnisse gefunden werden, die exakt diese Formulierung „*Apfel NEAR Birne*“ beinhalten, was nicht unbedingt auf eine Unterstützung des Operators schließen lässt. Ebenso unbestimmbar ist die Unterstützung des NEAR-Operators durch FARR, da die ersten vier der insgesamt 20 Suchergebnisse zwar „*Apfel und Birne*“ enthielten, die restlichen Ergebnisse aber keine eindeutige Relevanz bezüglich der Suchanfrage aufwiesen.

Aufgrund der Beschränkung der Zahl indexierbarer Dateien in der Testversion von Archivarius wurden nicht ausreichend viele Dateien indexiert, die die erwähnten Suchbegriffe enthalten. Für diese Suchmaschine wurde der Test stattdessen mit den Suchtermen „*Wissensmanagement*“ und „*Stelzer*“ durchgeführt. Eine vorherige Prüfung ergab, dass hierfür eine ausreichende Anzahl Dateien indexiert wurde.

Für Hulbee Desktop wird im zugehörigen Handbuch erläutert, dass alle Booleschen Operatoren AND, OR und NOT unterstützt werden. Jedoch zeigt der Test dass nur AND und OR akzeptiert wurden, denn für die Anfrage „*Apfel NOT Birne*“ wurden auch Suchergebnisse mit beiden Begriffen ausgegeben.

Standardverknüpfung

Im Bezug auf den vorausgegangenen Abschnitt stellt sich nun die Frage, welcher der genannten Operatoren standardmäßig verwendet wird, wenn ein Benutzer eine Suchanfrage aus zwei oder mehr Begriffen stellt und er diese nicht mit einem Operator verknüpft.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
Standardverknüpfung	AND	AND	AND	nicht eindeutig	OR	nicht eindeutig	OR

Tab. 10: Standardverknüpfung

Bei Windows, Google und Copernic wurde als Standardverknüpfung „AND“ registriert, da bei der Suche nach „*Apfel Birne*“ nur Treffer angezeigt werden, die beide Begriffe beinhalten. Archivarius zeigt hingegen neben Suchergebnissen mit beiden Begriffen auch welche mit nur „*Apfel*“ oder nur „*Birne*“ an, was auf die Verwendung von OR als Standardverknüpfung schließen lässt.

Sowohl bei xfriend als auch bei Hulbee wird im zugehörigen Handbuch die Verwendung der Verknüpfung mit „OR“ erwähnt. Allerdings werden bei Eingabe der Suchanfrage von beiden Programmen nur Ergebnisse geliefert, die beide Suchbegriffe beinhalten. Folglich könnte man im vorliegenden Fall von einer Standardverknüpfung mit AND ausgehen, was jedoch den Anmerkungen im Handbuch widerspricht.⁵² Dieser Widerspruch wird durch die „nicht eindeutig“ Kennzeichnung illustriert.

Bei FARR verhält es sich so, dass nur Treffer mit beiden Suchbegriffen oder mit „Apfel“ angezeigt werden, jedoch niemals nur „Birne“. Ändert man die Reihenfolge der Suchanfrage in „Birne Apfel“, so ändert das nichts am Ergebnis. Generell spricht dieser Sachverhalt dennoch dafür, dass FARR die OR-Verknüpfung verwendet.

Die Web-Suchmaschine Google verwendet standardmäßig den AND-Operator,⁵³ was zur Folge hat, dass viele Desktop-Suchmaschinen-Benutzer dieses Verhalten von einer Suchmaschine gewohnt sind. Aus diesem Grund sollen diejenigen Programme in der abschließenden Auswertung der Testergebnisse eine bessere Bewertung erhalten, die ebenfalls die AND-Verknüpfung als Standard verwenden.

Trunkierung

Für den Test, ob die Suchmaschinen Trunkierung unterstützen, und ob dabei die Position des Platzhalters am Anfang oder am Ende des gesuchten Begriffes ausgewertet wird, werden als Beispiel die Suchterme „*anleitung“ und „Erdbeer*“ verwendet.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
*anleitung	+	-	+/-	-	+	-	-
Erdbeer*	+	-	+	+	+	+	-
Σ	2	0	1,5	1	2	1	0

Tab. 11: Trunkierungsoperatoren

Beide Suchanfragen mit Trunkierungsoperator werden nur von Windows und Archivarius vollständig unterstützt. Windows findet alle zu den Anfragen passenden Begriffe, mit welchen die Testdaten vor der Untersuchung manipuliert wurden. Für *anleitung werden Dateien gefunden, die *Anleitung*, *Bedienungsanleitung*, *Reparaturanleitung* und *Anlei-*

⁵² Es wurde vor dem Test sichergestellt, dass Dateien mit nur einem der beide Suchbegriffe indiziert wurden. Ein Auffinden dieser durch die Suchmaschinen ist also generell möglich.

⁵³ Vgl. <http://www.google.com/support/websearch/bin/answer.py?answer=136861#exceptions>, Abruf: 2010-06-22

tungssammlung beinhalten. Für den Suchbegriff *Erdbeer** werden ebenso alle relevanten Ergebnisse, wie *Erdbeeren*, *Erdbeersalat* und *Erdbeerernte*, gefunden.

Archivarius weist gegenüber Windows den Vorteil auf, dass die gefundenen Begriffe in der Suchergebnisliste farbig hervorgehoben werden, wobei jeder Wortkombination eine bestimmte Farbe zugewiesen wird.

Xfriend und Hulbee akzeptieren nur die zweite Suchanfrage mit Platzhalter am Ende des Wortes, wobei Xfriend durch die Anzeige der Meldung "Ihre Suche darf nicht mit * oder ? beginnen" den Benutzer darauf aufmerksam macht. Im Handbuch von Hulbee Desktop wird erwähnt, dass mehrere Zeichen durch das Symbol „*“ ersetzt werden können. Jedoch gibt es hier keinen Hinweis darauf, dass das Sternchen nur am Ende des Suchbegriffs gesetzt werden kann.

Copernic erhält für das auf die Anfrage „*anleitung“ hin gelieferte Ergebnis die Bewertung „+/-“ (s. Tab. 11), da nicht alle möglichen Begriffe gefunden werden. Google und FARR unterstützen die Trunkierung nicht.

Speichern einer Suchanfrage

Hinsichtlich der Speicherung von einmal gestellten Suchanfragen lässt sich festhalten, dass sechs der sieben Desktop-Suchmaschinen diese Funktion unterstützen, wobei Find and run robot als einzige keine Speicherung vornimmt. Für Benutzer der anderen Suchmaschinen äußert sich die Unterstützung der Speicherung über die Anzeige vergangener Suchbegriffe in einer Dropdown-Liste unterhalb des Suchfelds, wobei die Darstellung bei jeder Suchmaschine etwas anders ausfällt.

Xfriend verfährt so, dass nach Eintippen eines Buchstabens alle Suchbegriffe in alphabetischer Reihenfolge angezeigt werden, die bisher zu diesem Anfangsbuchstaben gestellt wurden. Tippt man einen weiteren Buchstaben ein, so reduziert sich die Liste auf alle bisherigen Anfragen, die mit diesen beiden Buchstaben beginnen. Bestehen noch keine Anfragen zu dieser Kombination, dann zeigt xfriend auch frühere Suchbegriffe mit der Buchstabenkombination in der Wortmitte an, z.B. nach Eintippen von „Al“ wird „FußbALLspiel“ eingeblendet. Hulbee, Google und Windows zeigen ebenso automatisch frühere Suchanfragen mit gleichem Anfangsbuchstaben während der Eingabe.

Anders verhält es sich bei Archivarius und Copernic, bei denen der Benutzer zunächst eine Schaltfläche neben dem Suchfeld anklicken muss, damit die bisherigen Suchanfragen in

chronologischer Reihenfolge angezeigt werden. Hulbee verfügt zusätzlich auch über diese Funktion. Bei Google Desktop fällt auf, dass auch bereits in der Google Internetsuche oder in der Suchfunktion des Social Networking-Anbieters Facebook gestellte Suchanfragen gelistet werden.

Mit Ausnahme von Hulbee speichern alle Desktop-Suchmaschinen die Suchergebnisse über einen längeren Zeitraum, d.h. sie sind auch nach dem Neustart des PC noch verfügbar.

Inkrementelle Suche

Das Anzeigen von Suchergebnissen bereits während der Eingabe wird von FARR und Windows unterstützt, sowie teilweise von Google. Für FARR ist dieses Verhalten eine grundlegende Programmfunktion. Das Betätigen der Enter-Taste führt hier nicht, wie bei den anderen Suchmaschinen, zur Bestätigung der Suchanfrage, sondern dazu, dass die erste Datei der Suchergebnisliste geöffnet wird.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
Inkrementelle Suche	+	+/-	-	-	-	-	+

Tab. 12: Inkrementelle Suche

Auch bei Verwendung der Windows-Suche werden die in der Ergebnisliste angezeigten Dateien nach Eintippen jedes weiteren Buchstaben des Suchbegriffs aktualisiert.

Google Desktop Search zeigt die Dateinamen der Vorschläge während des Tippvorgangs in einem Dropdown-Menü an. Damit eine komplette Suchergebnisliste im vorgesehenen Bereich angezeigt wird, ist aber die Bestätigung mittels „Enter“ notwendig. Nicht unterstützt wird die Funktion von Copernic, Archivarius, xfriend und Hulbee. Copernic bietet die Suche bereits beim Tippen in den „erweiterten Optionen“ des Programms grundsätzlich an, jedoch kann sie nur ausgewählt werden, wenn die kostenpflichtigen Versionen Professional und Corporate verwendet werden. Der Benutzer wird, sobald er versucht die Option auszuwählen, durch einen Hinweis im Programm auf diesen Sachverhalt aufmerksam gemacht.

Durchsuchter Bereich

Mit Ausnahme von FARR durchsuchen alle Programme sowohl Dateiname als auch Dateiinhalt. Zur Überprüfung dieses Sachverhalts wird ein Suchbegriff festgelegt, der dem Inhaltstext verschiedener Dokumente hinzugefügt wird. Werden diese auf die Suchanfrage

hin gefunden, ist dies ein Hinweis darauf, dass auch der Inhalt geprüft wird. Genaue Festlegungen können für die Windows-Suche getroffen werden, da hier die Möglichkeit besteht, in den Indexierungsoptionen festzulegen, für welche Dateitypen nur die Eigenschaften und für welche ebenfalls die Dateiinhalte indexiert werden sollen. Meistens ist standardmäßig letztere Einstellung gesetzt.

Betrachtet man die Suchergebnisliste von Archivarius, so fällt als Besonderheit dieser Suchmaschine auf, dass für Audiodateien neben der Anzeige von Titel und Interpret auch die zusätzlich verfügbaren Metadaten eingeblendet werden, wie beispielsweise Albumtitel, Erscheinungsjahr oder Komponist. Diese können ebenfalls als Suchterme zum Filtern der Audiodateien verwendet werden.

6.7 Erweiterte Suchmöglichkeiten

Gegenstand der folgenden Untersuchung ist der Umfang, in welchem die Suchmaschinen dem Benutzer eine Möglichkeit bieten, die Suchanfrage zu verfeinern bzw. eine Filterung der Ergebnisse vorzunehmen.

Die von der Google-Internetsuche bekannte Sucheingrenzung „Ergebnisse finden...“⁵⁴, wird von der Google Desktop-Suche, sowie Xfriend und Archivarius unterstützt. Eine Eingrenzung der Ergebnisse auf eine bestimmte Sprache erlauben beide Suchmaschinen vor der Anfrage, während Hulbee diese erst danach erlaubt. Durch Klick auf eine Schaltfläche werden nur noch die Dateien in der ausgewählten Sprache angezeigt.⁵⁵

⁵⁴ Es handelt sich hier um die folgenden Formulierungen: Ergebnisse finden „mit allen Wörtern: *Suchbegriff*“; „mit der genauen Wortgruppe: *Suchbegriff*“; „mit irgendeinem der Wörter: *Suchbegriff*“; „ohne die Wörter: *Suchbegriff*“.

⁵⁵ Die in Anhang B verfügbare Abbildung der Hulbee-Benutzeroberfläche illustriert dies. Filtermöglichkeit.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
„Ergebnisse finden...“	-	+	-	+	+	-	-
Sprache	-	-	-	+	+	+	-
Dateiformat	+	-	+	+	+	+	-
Datum	+	+	+	+	+	+	-
Position des Suchbegriffs	+/-	+/-	-	-	-	-	-
Σ	2,5	2,5	2	4	4	3	0

Tab. 13: Erweiterte Suchmöglichkeiten

Eine Filterung nach Dateiformat bieten fünf der sieben Suchmaschinen an, wobei für Windows über die Option „Gruppieren nach: Typ“ nur eine Ranking-Möglichkeit zur Verfügung steht, bei welcher nach wie vor alle Suchergebnisse angezeigt werden und sich allein die Rangfolge ändert. Von insgesamt sechs Programmen wird die Filterung nach Datum unterstützt. Unklar bleibt dabei aber regelmäßig, ob mit dem Begriff *Datum* das Erstellungsdatum der Datei oder das letzte Veränderungsdatum gemeint ist. Einzig Archivarius verwendet eindeutig den Begriff des letzten Dateimodifizierungsdatums.

Die Position des Suchbegriffs näher zu bestimmen, ist bei keiner der getesteten Suchmaschinen ausdrücklich möglich. Für die Windows-Suchfunktion und Google besteht aber eine Option, mit welcher die Suchergebnisse nach Suchrelevanz angezeigt werden, was dazu führt, dass i.d.R. zuerst die Ergebnisse mit dem Suchbegriff im Dateinamen angezeigt werden.

6.8 Retrievalmaße

Gesamttrefferzahl

In Tabelle 14 sind die insgesamt von den einzelnen Desktop-Suchmaschinen für einen bestimmten Testsuchbegriff⁵⁶ ausgegebenen Trefferzahlen zusammengestellt.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
Gesamt-trefferzahl	26	153	195	146	%	176	65

Tab. 14: Gesamttrefferzahl

Für Archivarius sind aufgrund der in Kapitel 6.1 erwähnten Beschränkung der indexierbaren Dateien auf ein Maximum von 10000 Stück keine sinnvolle Messung und

⁵⁶ Suchbegriff: „Wissensmanagement“

somit auch kein Ranking im Vergleich mit den anderen Suchmaschinen möglich. Kritisch bleibt darüber hinaus die Bewertung der Gesamttrefferzahlen. Werden viele Ergebnisse angezeigt, so können diese zum Großteil relevant sein, aber ebenso ist es möglich, dass es sich um eine durch Ballast ausgeweitete Treffermenge handelt. Da üblicherweise jedoch eine größere Ergebnismenge auch eine größere Anzahl relevanter Dokumente beinhaltet [GrFr04, 4], wird das Ergebnis von Copernic mit der höchsten Gesamttrefferzahl von 195 Suchergebnissen am besten bewertet.

Precision (first ten precision)

Mit dem Maß Precision wird die Anzahl der relevanten Dokumente unter den ersten zehn Suchergebnissen ermittelt. Als mäßig relevant (s. Tab. 15) sollen im Folgenden diejenigen Dateien angesehen werden, in welchen der Testbegriff „Wissensmanagement“ zwar Erwähnung findet, aber im Sinne von Pertinenz, dem Befriedigen eines Informationsbedarfs, diesen nicht hundertprozentig erfüllt. Dies soll im Folgenden der Fall sein, wenn das *Wissensmanagement* nicht Hauptthema des Inhalts ist, sondern nur ein Gliederungspunkt eines komplexeren Hauptthemas. Als irrelevant werden solche Dateien gekennzeichnet, die den Term *Wissensmanagement* in keiner Weise beinhalten bzw. es zwar beinhalten, aber überhaupt keinen relevanten Beitrag zur Befriedigung des Informationsbedarfs leisten. Das kann beispielsweise bei einer persönlichen E-Mail der Fall sein, in welcher allein in der Signatur der Titel „Mitarbeiter des Fachgebiets Informations- und Wissensmanagement“ vermerkt ist, und der Inhalt ansonsten keinen Bezug zum Thema aufweist. Befinden sich unter den Suchergebnissen auch Dateiformate, deren Inhalt nicht-textuell ist, so wird das Ergebnis als relevant betrachtet, wenn der Suchbegriff im Dateinamen vorkommt.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
Relevante Ergebnisse	10	9	5*	10	7*	2	10
Mäßig relevante Ergebnisse	0	0	4*	0	3*	5	0
Irrelevante Ergebnisse	0	1	1*	0	0*	3	0

Tab. 15: Precision⁵⁷

Da mittels Copernic Desktop Search keine Suche über alle Dateien hinweg möglich ist, sondern immer automatisch eine Aufteilung in E-Mails, Dateien, Musik, Bilder und Videos erfolgt, ist das Ergebnis an dieser Stelle nicht vergleichbar. Als Ersatzwerte werden die auf Ebene der Dateien ausgegebenen Suchergebnisse in der Tabelle angegeben. Ebenso ist auch bei Archivarius, aufgrund der bereits bekannten Beschränkung des Index, der Vergleich schwierig. Hier wird ersatzweise auf die für die Suche im Verzeichnis C:\Benutzer\Benutzername\Eigene Dateien\DB\DB angezeigten Ergebnisse Bezug genommen. Die nicht vergleichbaren Werte sind in der Tabelle 15 mit einem Sternchen gekennzeichnet.

Relativer Recall

Problematisch werden die nicht vergleichbaren Ersatzwerte wenn es nun um die Berechnung des relativen Recall geht. Bildet man die Summe aller relevanten Treffer, um dann den Anteil jeder einzelnen Suchmaschine zu ermitteln, würde es zur Verfälschung der Ergebnisse kommen. Folglich müssen Ersatzwerte gefunden werden, die einen Vergleich dennoch möglich machen. Eine Option ist das arithmetische Mittel:

$$\bar{x}_{\text{arithm}} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Da es auch für die Berechnung von Ersatzwerten für die Gesamtauswertung von Bedeutung ist, sei an dieser Stelle auf die weiteren Erläuterungen in Kapitel sieben verwiesen.

⁵⁷ Die in der Tabelle angegebenen Werte beschreiben, wie viele Dateien innerhalb der ersten zehn Suchergebnisse einer Suchmaschine, den jeweiligen Ergebniskategorien (relevante, mäßig relevante und irrelevante Ergebnisse) zugehören. Bei den mit Sternchen-Symbol gekennzeichneten Werten handelt es sich um Ersatzwerte.

Ersetzt man also die für Copernic und Archivarius in der Tabelle 16 angegebenen Werte durch den berechneten Mittelwert (8,2), so erhält man für alle Suchmaschinen eine Summe von 57,4 relevanten Ergebnissen (s. Tab. 16).

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR	Σ
Relevante Ergebnisse	10	9	8,2	10	8,2	2	10	57,4

Tab. 16: Relevante Ergebnisse mit Ersatzwerten⁵⁸

Mit den Werten der Tabelle 16 ergibt sich den relativen Recall folgende Verteilung:

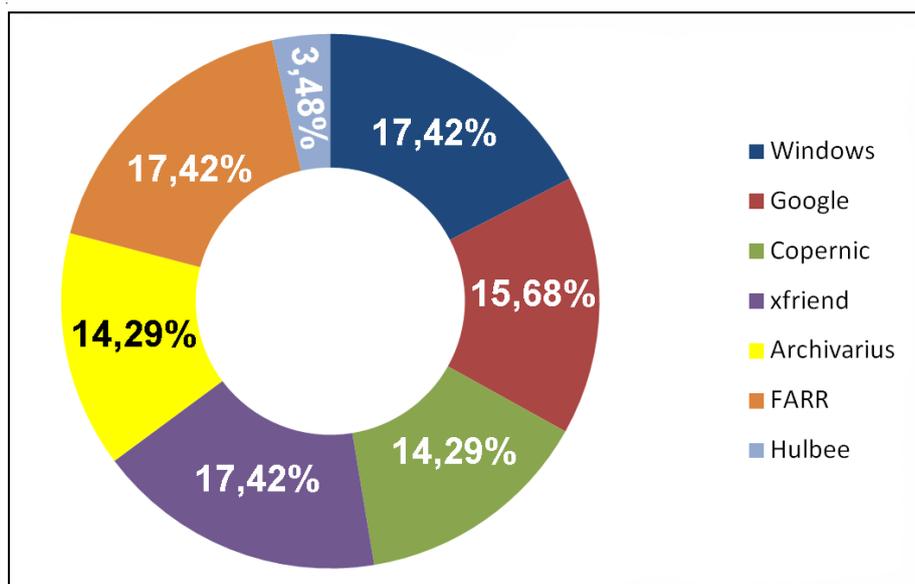


Abb. 5: Relativer Recall

Betrachtet man das in der Abbildung 5 dargestellte Tortendiagramm, so wird deutlich, dass die Anteile der einzelnen Suchmaschinen an den von allen Suchmaschinen ausgegebenen Treffern relativ gleichmäßig verteilt sind. Grund hierfür ist, dass die aus dem Mittelwert berechneten relativen Werte für Copernic und Archivarius, bei der gegebenen Ausgangssituation von drei Suchmaschinen mit dem Maximum an relevanten Ergebnissen, einer mit der zweitgrößten möglichen Trefferzahl und nur einem negativen Ausreißer, verhältnismäßig hoch ausfallen.

⁵⁸ Die Zeile „Relevante Ergebnisse“ wurde aus Tab. 15 übernommen.
Die rot dargestellten Werte entsprechen dem arithmetischen Mittel der restlichen Einzelwerte.

6.9 Indexierung

In den folgenden Abschnitten wird die Indexqualität der Desktop-Suchmaschinen getestet, wobei auf FARR nicht eingegangen wird, da das Programm, wie bereits bekannt ist, keinen Index erstellt.

Die vom Benutzer individuell konfigurierbare Index-Qualität von xfriend muss an dieser Stelle besonders erwähnt werden. Die Abstufungen „normal“, „sehr gut“ und „maximal“ richten sich nach der Anzahl der indexierten Wörter eines Dokuments, wobei die normale Index-Qualität voreingestellt ist. Ziel ist jedoch auch in diesem Punkt eine Bewertung der einzelnen Desktop-Suchmaschinen auf gleicher Ausgangsbasis. Hierzu werden die nachfolgenden Kriterien, wie Beeinflussungsmöglichkeit oder Aktualität des Index für jede Suchmaschine ermittelt.

Beeinflussungsmöglichkeiten des Benutzers

In der nachfolgenden Tabelle ist dargestellt, inwieweit der Benutzer bei den einzelnen Desktop-Suchmaschinen Einfluss auf den Indexierungsvorgang nehmen kann.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
Dateiformate	+	+	+	+	+	-	%
Ordner & Laufwerke	+	+	+	+	+	-	%
Zeitpunkt	+/-	-	+	+	+	-	%
Σ	2,5	2	3	3	3	0	%

Tab. 17: Beeinflussung der Indexierung durch den Benutzer

Mit Ausnahme von Hulbee bieten alle Suchmaschinen die Möglichkeit, eine Auswahl hinsichtlich der zu indexierenden Dateiformate zu treffen. I.d.R. ist eine Liste von Dateitypen voreingestellt, aus welcher entsprechend der persönlichen Präferenzen diejenigen entfernt werden können, die nicht in den Index aufgenommen werden sollen.

Die Vorauswahl bzgl. zu indexierender Verzeichnisse und Ordner wird ebenfalls von allen Suchmaschinen außer Hulbee unterstützt. Google weist in den Einstellungen darauf hin, dass alle Festlaufwerke standardmäßig indexiert werden. Zusätzliche Laufwerke können hinzugefügt werden. Bei den vier anderen, untersuchten Programmen, die diese Option unterstützen, können Unterschiede hinsichtlich des Umfangs der Auswahl festgestellt werden. Copernic ermöglicht nur die Auswahl der einzelnen Laufwerke, sowie des Desktops und der Ordner „Eigene Dokumente“, „Eigene Bilder“, „Eigene Musik“ und „Eigene

Videos“, wohingegen bei Windows, xfriend und Archivarius die Auswahl jedes einzelnen Unterordners über die komplette Gliederungstiefe möglich ist.

Der Zeitpunkt der Indexierung kann für Archivarius, xfriend und Copernic individuell festgelegt werden. Der Benutzer von Archivarius kann hierbei sowohl die Startzeit wählen als auch die Häufigkeit der Indexierung. Ebenso kann eingestellt werden, ob bei jedem Indexierungsvorgang nur veränderte, neue oder alle Dateien indexiert werden. Interessant und bedenkenswert ist, dass in den Grundeinstellungen von Archivarius eine einmalige Indexierung voreingestellt ist. Archivarius würde also, ohne Veränderung der Einstellung durch den Benutzer, nur einmalig nach der Erstinstallation indexieren, was zur Folge hätte, dass neu hinzukommende Dateien nicht gefunden werden können.

Bei Copernic kann der Nutzer, abweichend von der Standardeinstellung, den Zeitpunkt und die Häufigkeit der Indexierung selbst festlegen. Die Häufigkeit kann zwischen „alle Tage“ und „n Tage“ variieren. Es besteht zudem die Möglichkeit für verschiedene Quellen unterschiedliche Aktualisierungsrhythmen festzulegen. So kann ein Anwender beispielsweise täglich Bilddateien aktualisieren, Textdokumente hingegen nur alle drei Tage. Eine komplette Deaktivierung der Indexierung für eine Quelle ist ebenso möglich.

Xfriend bietet sowohl eine manuelle als auch eine automatische Indexierung an, wobei der mögliche festlegbare Zeitraum für letztere sich von „eine Minute“ bis auf „n Tage“ erstreckt. Mit Hilfe einer kalendarischen Indexierung kann festgelegt werden, zu welchem Kalenderdatum die nächste Indexierung ausgeführt werden soll.

Bei der Windows-Suche können nur die Einstellungen „automatisch“, „manuell“ und „deaktiviert“ gesetzt werden. Genauere Festlegungen, die im Zusammenhang mit den anderen Programmen erläutert wurden, können nicht getroffen werden.

Aktualität des Index

Zur Messung der Aktualität wird unmittelbar nach dem Abspeichern eines neuerstellten Word-Dokuments die erste Suchanfrage gestellt, welche sich am Dateinamen des Dokuments orientiert. Wird das Dokument nicht gefunden, so erfolgt eine minütliche Wiederholung der Anfrage, bis es erstmals angezeigt wird.

Copernic und die Windows-Suche finden das Dokument nach der ersten Anfrage. Für xfriend und Archivarius ist ein ebenso schnelles Auffinden möglich, vorausgesetzt es

wurde vom Benutzer eine minütliche Aktualisierung des Index festgelegt. Google Desktop Search findet die neuerstellte Datei nach elf und Hulbee nach 22 Minuten.

Indexierungsdauer

Startzeitpunkt für die Messung der Indexierungsdauer ist entweder der unmittelbare Abschluss des Installationsvorganges oder, soweit vorhanden, die Meldung über den Beginn der Indexierung. Endzeitpunkt ist folglich die Meldung über das Ende der Indexierung oder, wenn diese nicht angezeigt wird, das erstmalige Abfallen der CPU-Auslastung auf eine für den normalen Betriebszustand übliche Belastung.⁵⁹

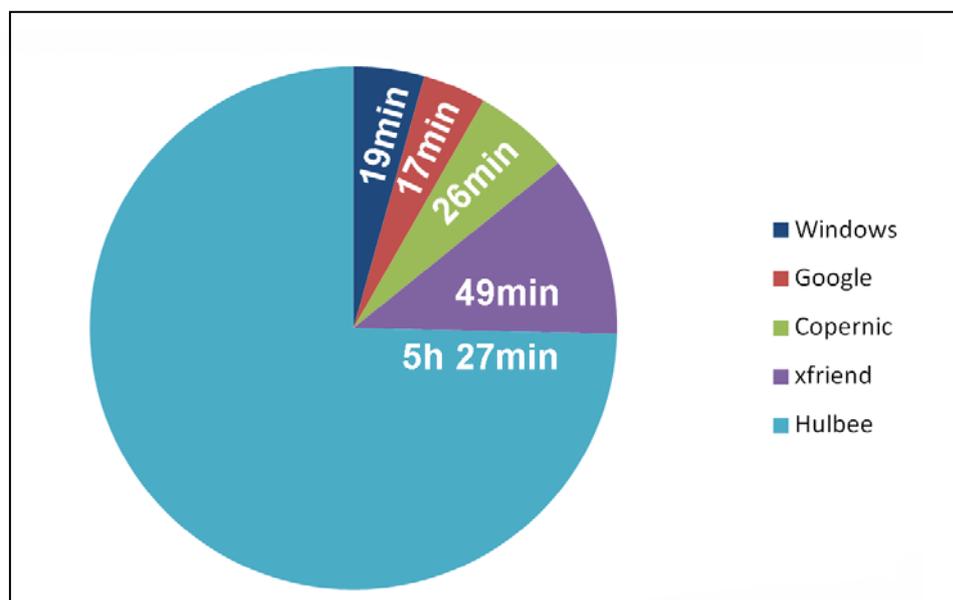


Abb. 6: Indexierungsdauer

Die im Diagramm der Abbildung 6 dargestellten Werte zeigen, dass hinsichtlich der Indexierungsdauer große Unterschiede zwischen den einzelnen Suchmaschinen bestehen. So liegen zwischen dem besten und dem schlechtesten Ergebnis insgesamt fünf Stunden und zehn Minuten. Google Desktop Search benötigt für die Indexierung aller Testdateien 17 Minuten, wohingegen Hulbee fünf Stunden und 27 Minuten indexiert.⁶⁰ Copernic benötigt für die Indexierung 23 Minuten unter Verwendung aller Computerressourcen und 26 Minuten in der für den geringeren Ressourcenbedarf optimierten Indexierungseinstellung. Für

⁵⁹ Dieser Übergang ist offensichtlich erkennbar, da während der gesamten Indexierung ausgeprägte Schwankungen in sehr kurzen Intervallen auftreten (s. hierzu auch Kapitel 6.2 → Speicherauslastung).

⁶⁰ Auszuschließen ist eine Benachteiligung von Hulbee aufgrund einer im Vergleich extrem größer ausfallenden Dateimenge, die indexiert werden muss. Vgl. hierzu die Ausführungen in Anhang A zur Indexierung des kompletten Laufwerks C: und zu nicht vorhandenen externen Laufwerken.

Archivarius ist aus den bereits erläuterten Gründen kein Vergleich mit den anderen Desktop-Suchmaschinen möglich.

Indexgröße

Abbildung 7 zeigt den Vergleich der Suchmaschinen im Hinblick auf die Indexgröße.⁶¹ Ein Einbeziehen von Archivarius ist in diesem Fall wiederum nicht möglich.

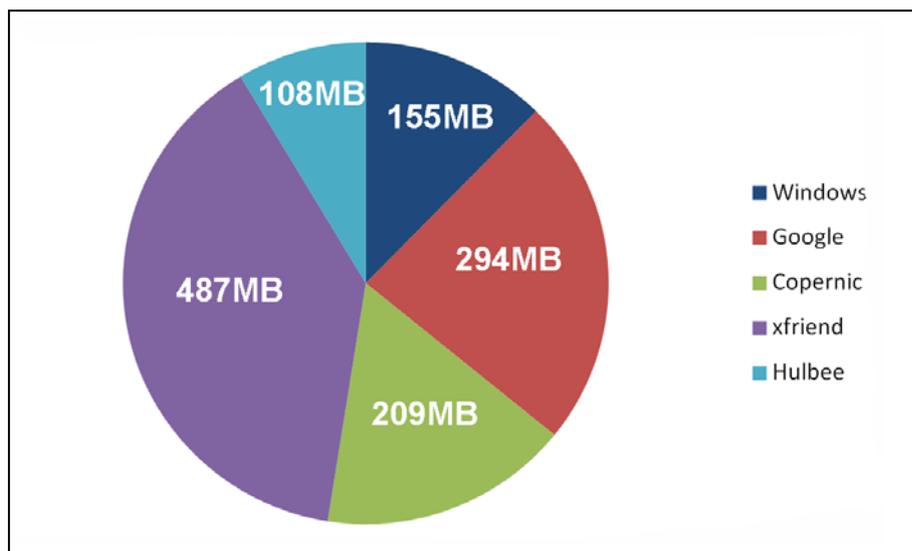


Abb. 7: Indexgröße

Wie aus Abbildung 7 ersichtlich, sind die Unterschiede zwischen den Resultaten der einzelnen Suchmaschinen bzgl. der Indexgröße geringer als im Vergleich mit den Werten der Indexierungsdauer betrachtet. Ein extremer Ausreißer gibt es hier nicht, auch wenn der größte Wert, die Indexgröße von xfriend mit 487 MB, mehr als viermal so groß ist wie der kleinste Wert, die für Hulbee gemessene Indexgröße von 108 MB. Dieser kompakte Index wird allerdings mit der längsten Indexierungszeit ‚erkaufte‘ (s. Abbildung 6). Von einer direkten Abhängigkeit der beiden Kriterien ist somit nicht auszugehen.

6.10 Anpassungsfähigkeit

Darstellungsformen

In Tabelle 18 sind alle vorkommenden Darstellungsformen der untersuchten Desktop-Suchmaschinen aufgelistet.

Die meisten Darstellungsformen bietet Google Desktop an. Der Benutzer hat die Möglich-

⁶¹ Grundlagen zur Messung sind dem Anhang E zu entnehmen.

keit auszuwählen, welche Elemente er verwenden möchte, da alle Optionen mit Ausnahme des browserbasierten Hauptfensters deaktiviert werden können. Die Entscheidung darüber, ob Google Desktop die für die Windows-Suche spezifischen Optionen anbietet, muss bereits während der Installation festgelegt werden. Entfernt man während dieses Vorgangs einen standardmäßig gesetzten Haken nicht aus der Check-Box für die sogenannte „erweiterte Suche“ mit der Anmerkung „Für diese Funktion wird der Google Desktop-Index anstatt des Vista-Indexes verwendet [...]“, wird die Windows-Suche, wie in der Beschreibung angegeben, automatisch deaktiviert und durch die Google-Suche ersetzt.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
Eigenständiges Programm	-	-	+	-	+	+	+
Browser-integration	-	+	-	+	-	-	-
Seitenleiste	-	+	-	-	-	-	-
Deskbar	-	+	+	-	-	-	-
Floating Deskbar	-	+	-	+	-	+	-
Schnellsuchfeld	-	+	-	-	-	-	+
Startmenü	+	+	-	-	-	-	-
Windows-Explorer	+	+	-	-	-	-	-
Browser-Toolbar	-	+	+	-	-	+	-
Σ	2	8	3	2	1	3	2

Tab. 18: Darstellungsformen

Bei der für Find and run robot bestätigten Schnellsuche, handelt es sich um exakt die gleiche Darstellungsform wie beim selbstständigen Programm. Man kann sagen, dass das Tool mit einem Schnellsuchfeld gleichgesetzt werden kann, da seine Oberfläche nur einen Bruchteil des Bildschirms einnimmt, es über Tastenkombinationen aufgerufen werden kann und ebenfalls die Funktion zum Programmöffnen unterstützt.

Unterstützte Dateiformate

Archivarius informiert den Benutzer während der Benutzung des Indexassistenten darüber, welche Dateitypen standardmäßig indexiert werden. Es handelt sich hierbei um ca. 250 Formate. Xfriend stellt eine ausführliche PDF-Datei zur Verfügung, die zusammen mit der Installationsdatei heruntergeladen wird. Hier werden alle unterstützten Dateiformate gelis-

tet. Zusätzlich werden Informationen zu jedem Format gegeben, beispielsweise ob eine Vorschaufunktion verfügbar ist und welche Metadaten indexiert werden.⁶²

Die meisten Suchmaschinen bieten eine derartige Auflistung jedoch nicht an. Da unter diesen Umständen nur schwer Aussagen getroffen werden können, welche Programme die meisten Dateiformate unterstützen, werden die in Kapitel zwei genannten Formate als Ausgangsbasis für den Test verwendet.⁶³

Das Ergebnis zeigt, dass außer Hulbee alle Suchmaschinen den gesamten Katalog an Dateiformaten unterstützen. Hulbee findet mit .rar, .zip, .tar und .iso vier der genannten Formate nicht, was eine geringfügig schlechtere Gesamtbewertung zur Folge hat.

Unterstützte spezielle Anwendungen

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
E-Mails	+	+	+	+	+	+	-
Browser/ Websuche	+	+	+	+	-	+	+
Instant- Messenger	-	+	-	-	+	-	-
Σ	2	3	2	2	2	2	1

Tab. 19: Spezielle Anwendungen

Das Durchsuchen des Inhalts von E-Mails ermöglichen sechs der sieben getesteten Suchmaschinen, wobei hier Unterschiede bzgl. des Umfangs bestehen. Windows und Hulbee unterstützen nur Microsoft Outlook. Weitere E-Mail-Clients wie Vista Mail, Mozilla Thunderbird und Lotus Notes werden von Hulbee nur in der kostenpflichtigen Professional-Version unterstützt.⁶⁴ Xfriend verfügt über die Option, dass im Installationsassistenten auch direkt eine Verbindung zu jedem beliebigen E-Mail-Konto angegeben werden kann. Die Verwendung eines E-Mail-Clients wie Outlook oder Thunderbird ist hier für die Indexierung der E-Mails infolgedessen nicht vorausgesetzt. Während bei Google auch der Inhalt der E-Mails durchsucht werden kann, bietet Copernic nur die Suche in der Betreffzeile, dafür jedoch viele weitere Optionen, wie die Suche nach dem Absender, dem Datum

⁶² Ein Auszug dieser Liste kann Anhang F entnommen werden.

⁶³ Details zum Messvorgang sind Anhang G zu entnehmen.

⁶⁴ Vgl. http://company.hulbee.com/products_and_solutions/hulbee_desktop_pro/, Abruf: 2010-06-22

oder der Wichtigkeit an. Mit insgesamt zwölf zur Auswahl stehenden Programmen unterstützt Archivarius die meisten E-Mail-Clients.

Die Möglichkeit, neben der Desktop-Suche auch im Internet zu suchen, bieten ebenfalls sechs der getesteten Programme an. Bei Google, Hulbee, Windows, Copernic und xfriend kann die Web-Suche direkt über die Auswahl dieser Option im Programm durchgeführt werden. FARR führt eine Suche im Internet durch wenn man eine Suchanfrage mit dem Zusatz „search“ eingibt, also beispielsweise „search Informationsmanagement“. Als Suchergebnis erscheinen daraufhin Links zu verschiedenen Suchmaschinen wie Google, Yahoo oder Ask. Durch Klick auf einen der Verweise wird die Suchergebnisseite der jeweiligen Suchmaschine angezeigt.

Die Unterstützung der Suche nach Instant Messenger-Protokollen wird in der Testphase nicht geprüft. Jedoch verweisen Google und Archivarius auf die Unterstützung dieser Funktion,⁶⁵ was durch eine positive Bewertung in Tabelle 20 festgehalten wird.

Unterstützung gelöschter und gepackter Dateien

Wurde eine Datei in den Papierkorb verschoben, so wird sie bei einer entsprechenden Suchanfrage von den Suchmaschinen Windows, Copernic, Hulbee und FARR nicht mehr in der Ergebnisliste angezeigt. Archivarius, xfriend und Google zeigen eine Datei auch noch an, wenn sie bereits in den Papierkorb verschoben, wurde. Möchte man die Datei jedoch öffnen, so erhält der Benutzer eine Fehlermeldung, die verdeutlicht, dass die Datei nicht geöffnet bzw. gefunden werden kann. Einen Hinweis für den Anwender, in welchem die Verschiebung in den Papierkorb als Ursache erwähnt wird, gibt jedoch keine der drei Desktop-Suchmaschinen.

Wie aus Tabelle 20 ersichtlich, werden die in einer Archivdatei befindliche Dateien nur von Archivarius und xfriend gefunden, wobei letztere Suchmaschine bei Anfrage „Asien“ das im Archiv „urlaubsbilder.rar“ befindliche „asien.jpg“ sogar an erster Stelle der Suchergebnisliste anzeigt. Hulbee bietet die Funktion in der Professional-Version an.

⁶⁵ Vgl. <http://desktop.google.com/de/features.html>, <http://www.likasoft.com/de/order.shtml>,
Abruf :2010-06-22

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
Gelöschte Dateien	-	+	-	+	+	-	-
Gepackte Dateien	-	-	-	+	+	-	-
Σ	0	1	0	2	2	0	0

Tab. 20: Unterstützung gelöschter und gepackter Dateien

6.11 Vertraulichkeit

Standpunkt des Anbieters

Google, Copernic und xfriend verweisen jeweils in den Lizenzvereinbarungen während der Installation auf ihre Datenschutzrichtlinien. Durch einen Link erhält man auf den jeweiligen Webseiten detailliertere Information. Generell gestalten sich die Standpunkte der Anbieter aber ähnlich.

Für Google wird eine Einstellung standardmäßig gesetzt, die durchaus als kritisch betrachtet werden kann. Es handelt sich hier um die Option „Senden von Fehlerberichten und anonymen Nutzungsdaten“ um „Google Desktop zu verbessern“. Google gibt hierzu an, dass eine Information an das Unternehmen gesendet wird, sobald die Anwendung installiert oder entfernt wird. Außerdem werden Daten zum Betriebssystem des Nutzers, zur Anzahl der ausgeführten Suchvorgänge oder bzgl. der für das Anzeigen der Ergebnisse benötigten Zeit gesendet. Um persönliche Inhalte, wie personalisierte Nachrichten, bereitzustellen, können auch Informationen zu den vom Benutzer aufgerufenen Webseiten an Google weitergeleitet werden.⁶⁶ Google erklärt jedoch, keine Daten an Dritte weiterzugeben.

Die Datenschutzbestimmungen von Copernic stimmen in allen eben erläuterten Punkten mit denen von Google überein. Für xfriend wird hingegen ausdrücklich erwähnt, dass keinerlei Informationen über das Nutzungsverhalten gesammelt werden. Die anderen Aspekte sind ebenfalls identisch mit Google und Copernic.

Für Archivarius und Hulbee sind nur die für die Internetseite ausgelegten Datenschutzrichtlinien verfügbar. Softwarespezifische Richtlinien wurden nicht gefunden. Gleiches gilt für die Windows-Suche und FARR.

⁶⁶ Vgl. <http://desktop.google.de/de/privacyfaq.html>, Abruf: 2010-06-22.

Schutz vor unbefugten Zugriffen

Archivarius bietet als einzige Suchmaschine die Möglichkeit an, ein Passwort einzurichten, durch welches der Nutzer die Anwendung sperren kann und eine Desktop-Suche für Unbefugte somit schwer möglich macht.

Das Durchsuchen der Web-History bieten nur die Windows-Suche, Google und Copernic an. Einerseits stellt das natürlich einen Zusatznutzen für den Anwender dar, andererseits besteht aber die Gefahr, dass Dritte am Arbeitsplatz des Anwenders diese Funktion unbefugt nutzen können.

Als Schutz gegen solches Vorgehen können in Windows weitere Benutzerkonten eingerichtet werden, wenn auch Dritte den PC benutzen. Um zu testen, ob persönliche Daten des Hauptbenutzers nicht von einem solchen fremden Benutzerkonto aus gesucht bzw. geöffnet werden können, wurden zwei zusätzliche Konten, davon eines mit und eines ohne Admin-Rechte, eingerichtet. Es stellt sich heraus, dass die Suchmaschinen zwar zum größten Teil auch für diese Konten zur Verfügung stehen, jedoch legen sie einen eigenen Index an und verwenden nicht den des Hauptnutzers. Eine Desktop-Suche in den privaten Daten des Hauptbenutzers von einem Drittzugang aus ist folglich bei keiner Suchmaschine möglich.

6.12 Darstellung der Suchergebnisse

Der letzte Abschnitt zur Bewertung von Desktop-Suchmaschinen fasst alle Aspekte die Darstellungsweise der Suchergebnisse betreffend zusammen.

Anzeige des Dateinamens

Bei der Darstellung der Suchergebnisse steht der Dateiname im Mittelpunkt. Er wird i.d.R. durch eine große Schriftgröße, Fettschrift oder Unterstreichung hervorgehoben. Allein bei Copernic fällt auf, dass für das Container-Dateiformat Ogg der Dateiname nicht angezeigt wird, sondern stets nur der, in den Testdaten oft abweichende, Songtitel der Audiodatei.

Anzeige des Speicherortes

Bei xfriend, Google, Hulbee und Copernic wird der Speicherort grundsätzlich immer angezeigt. Die Windows-Suche zeigt in einigen Fällen den Pfad zur Datei nicht an, wobei jedoch kein eindeutiges Schema erkennbar ist. Bei Archivarius ist die Darstellung des Speicherortes nicht standardmäßig eingestellt, kann aber vom Anwender geändert werden.

Bei FARR ist unter Verwendung der Basiseinstellungen „one Level back“ nicht der komplette Pfad ersichtlich, da hier nur der Ordner angezeigt wird, in welchem sich die Datei unmittelbar befindet. Insgesamt werden vier Einstellungsmöglichkeiten angeboten, darunter auch die Anzeige des vollständigen Pfades („full path“).

Anzeige des Erstelldatums bzw. letzten Veränderungsdatums

Die Windows-Suchfunktion zeigt das Änderungsdatum nur für einige Dateiformate standardmäßig in den Suchergebnissen an, für MP3-Dateien beispielsweise wird stattdessen die Länge oder auch keine ersetzende Angabe gemacht. Copernic zeigt ebenso bei Musik- und Videodateien die Länge statt des Datums an, Für FARR muss die Datumsanzeige erst vom Benutzer ausgewählt werden, da sie nicht standardmäßig angezeigt wird.

Vorschau

Einige Zeilen aus dem Inhalt von Textdateien, sowie Thumbnails von Bild- und Videodateien werden von Windows und Google angezeigt. Xfriend zeigt ebenso Ausschnitte aus dem Textinhalt, jedoch werden Thumbnails nur für Dokumente optional angeboten. Dem Benutzer wird dann eine verkleinerte Ansicht des Dokuments angezeigt, während für Mediadateien der Hinweis „Keine Vorschau verfügbar“ erfolgt. Archivarius blendet den Inhalt der gerade angewählten Datei der Suchergebnisliste in einem im rechten Bildschirmbereich liegenden Vorschaufenster ein. Für Dokumente wird der komplette Text formatunabhängig in Plain Text-Darstellung angezeigt, was für den Benutzer sehr unübersichtlich wirkt.⁶⁷ Bilder erscheinen ebenso komplett und für Videos und Musikdateien werden die vorhandenen Metadaten im Vorschaufenster angezeigt.

Direktes Öffnen

Copernic Desktop Search und die Windows-Suche sind in der Lage, Dateien in ihrer jeweiligen Darstellungsform direkt im Programm zu öffnen, ohne dass dafür dritte Anwendungen gestartet werden müssen. Damit Windows diese Funktion unterstützt muss eine Änderung an der Basiseinstellung durchgeführt werden, sonst wird nur die im Abschnitt „Vorschau“ beschriebene Darstellung angezeigt.

PDF-Dateien können so beispielsweise in ihrem kompletten Layout dargestellt werden und Video- bzw. Musikdateien werden über einen integrierten Player sofort abgespielt.

⁶⁷ Vgl. hierzu die in Anhang B dargestellte Abbildung der Benutzeroberfläche von Archivarius.

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
Anzeige Dateiname	+	+	+/-	+	+	+	+
Anzeige Speicherort	+/-	+	+	+	+	+	+/-
Anzeige Datum	+/-	+	+/-	+	+	+	+/-
Anzeige Vorschau	+	+	+	+	+	-	-
Direktes Öffnen	+	-	+	-	-	-	-
Highlighting	+	+/-	+	+	+	-	-
Eindeutiges Ranking	+	+	+	+	-	+	+
Σ	6	5,5	6	6	5	4	3

Tab. 21: Darstellung der Suchergebnisse

Highlighting

Die farbliche Hervorhebung der Suchergebnisse wird von Windows, Copernic, xfriend und Archivarius durchgeführt, wobei letztgenannte Suchmaschine bei mehrgliedrigen Suchtermen jeden einzelnen Begriff in einer anderen Farbe unterlegt.

Bei Google werden die Suchbegriffe durch Fettschrift hervorgehoben. So wird die Suchergebnisliste für den Benutzer allerdings wesentlich übersichtlicher, als wenn überhaupt keine Hervorhebung vorgenommen wird, wie bei Hulbee und FARR.⁶⁸

Ranking

Das standardmäßig verwendete Rankingoption wird bei allen Suchmaschinen dem Benutzer gegenüber offen gelegt. Xfriend und Hulbee sortieren die Suchergebnisliste nach Relevanz, Copernic und Google sortieren nach dem Datum, die Windows-Suche nach dem Dateinamen. Archivarius erstellt Bewertungen der Suchergebnisse, welche jedoch nicht in jeder Einstellung sichtbar sind. Es wird allerdings nicht transparent gemacht, wie genau sie zustande kommen. Da Suchbegriffe i.d.R. häufiger in den höher gerankten Dateien auftauchen, kann vermutet werden, dass klassische, statistische Verfahren der automatischen Indexierung verwendet werden. Auch FARR verwendet ein Bewertungssystem, welches im Gegensatz zu den Mitbewerbern offenliegt und aktiv zu verändert werden kann. Eine von vielen Möglichkeiten besteht beispielsweise darin, bestimmten Dateitypen höher zu gewichten, um diese an den Beginn der Suchergebnisliste zu platzieren.

⁶⁸ Vgl. hierzu die in Anhang B dargestellten Abbildungen aller Suchmaschinen.

Archivarius und FARR, mit ihrem eigenen Bewertungssystem, bieten keine weiteren Rankingmöglichkeiten an. Google verfügt über eine weitere Option, die Auflistung nach Relevanz. Xfriend, Copernic und Hulbee bieten zwischen drei und sechs Möglichkeiten an und Windows gibt eine Liste von etwa 200 Kriterien an, nach denen die Ergebnisse sortiert werden können.

7 Auswertung

Dieses Kapitel fasst die im vorangegangenen Abschnitt ausführlich dargestellten Ergebnisse des Software-Tests zusammen, wobei die Bewertung zunächst untergliedert in die drei zu Beginn beschriebenen Untersuchungsebenen erfolgt. Auf diese Weise können Stärken und Schwächen der einzelnen Suchmaschinen in den unterschiedlichen Bereichen besser dargestellt werden als in den detaillierten Ergebnisbeschreibungen oder in der abschließenden Gesamtbewertung. Ein potentieller Nutzer ist so in der Lage zu erkennen, welches Programm im Hinblick auf seine persönlichen Präferenzen das beste Ergebnis erzielt. Grundlage der Bewertung sind die im 6. Kapitel vorgestellten Testergebnisse, welche, um einen Vergleich durchführen zu können, zunächst normiert werden, um eine Über- oder Untergewichtung einzelner Kriterien zu verhindern. Wenn keine abweichenden Anmerkungen vorhanden sind, wird, abhängig vom angestrebten Ziel, auf das Maximum oder das Minimum normiert.⁶⁹ Durch die Normierung können die Suchmaschinen in den nachfolgenden Tabellen für jedes Kriterium einheitlich mit den Punktzahlen 2, 1,5, 1, 0,5 und 0 bewertet werden.⁷⁰

Aufgrund verschiedener Faktoren, die an den betreffenden Stellen des fünften Kapitels erläutert wurden, konnte für einige Suchmaschinen hinsichtlich bestimmter Kriterien keine Messung durchgeführt werden. Um eine Verzerrung der Ergebnisse zu verhindern wird in den Tabellen zur Auswertung eine Imputation durchgeführt, was bedeutet, dass für die fehlenden Werte Ersatzwerte eingefügt werden [THNi04, 16 f.]. Als Ersatzwert wird, wie schon in Kapitel 6.8, das arithmetische Mittel für das jeweilige Kriterium verwendet. Zwar entstehen so Werte, die unmöglich sind, da FARR beispielsweise keinen Index anlegt und so eine Bewertung der Indexierung auch nicht möglich ist. Dennoch ist dieses Verfahren im Vergleich zu anderen Möglichkeiten zu bevorzugen.⁷¹ Die Ersatzwerte sind in den Tabellen rot hervorgehoben.

⁶⁹ Vgl. <http://www.husvedvf.de/index.php?view=article&id=128>, Abruf: 2010-06-21; Ein Beispiel zur Normierung ist Anhang H zu entnehmen.

⁷⁰ Das relativ niedrige Punktespektrum wird gewählt, da so eine Übergewichtung von Kriterien mit geringem Umfang (z.B. Tab. 17) gegenüber den aus vielen Einzelkriterien bestehenden (z.B. Tab. 22) vermieden wird.

⁷¹ Möglich wäre auch das Entfernen aller Kriterien bzw. Zeilen, in denen sich fehlende Werte befinden, was jedoch den Verlust wichtiger Information bedeuten würde. Andere Imputationsverfahren, z.B. Nearest Neighbor, garantieren zwar das Einsetzen von tatsächlich in der Stichprobe vorkommenden

7.1 Software-Ebene

								
	Kriterien	Copernic	Google	Windows	FARR	Hulbee	Archivarius	xfriend
6.1	Vielseitigkeit & Einschränkungen	1	2	2	1	1	0,5	1,5
6.2	Effizienz							
	Größe Installationsdatei	2	2	1,5	2	1	2	0
	Größe Programmordner	1,5	2	1,5	2	1,5	2	0
	Speicherauslastung							
	Programmbetrieb CPU	0,5	1,5	1,5	2	1	0	0,5
	Programmbetrieb RAM	0,5	0	0	0	0	0	0
	Suchvorgang CPU	1,5	1,5	1	0	0	0,5	1
	Suchvorgang RAM	0,5	0	0	0	0	0	0
	Indexierung CPU	2	0	1,5	1	1	0,5	1
	Indexierung RAM	0	0	0	0	0	0	0
6.3	Usability*	2	2	1,5	1	2	1	2
	Σ	11,5	11	10,5	9	7,5	6,5	6

Tab. 22: Bewertung der Testergebnisse auf Software-Ebene⁷²

Werten, jedoch würden sie das Ergebnis mehr verzerren als der Mittelwert, falls der beste oder schlechteste Wert einzusetzen wäre [THNi04, 13, 22]

⁷² Für die mit Sternchen gekennzeichneten Kapitelbezeichnungen in den nachfolgenden Tabellen 23 bis 25 gelten die in Anhang I dargestellten abweichenden Bewertungen.

7.2 Suchmaschinen-Ebene

								
	Kriterien	xfriend	Copernic	Windows	Google	Archivarius	Hulbee	FARR
6.4	Sprachliche Analyse							
	Stemming	0,5	0,5	0	0	0,5	0,5	1
	Singular und Plural	2	1	1	2	1	1	1,5
	Groß- und Kleinschreibung	0	0	0	0	0	0	0
	Soundex	0	0	0	0	0	0,5	1,5
6.5	Rechtschreibprüfung	2	2	0	0	0	0	1
6.6	Best. Suchfunktionen							
	Operatoren	2	2	1,5	0,5	0,5	1	1,5
	Standardverknüpfung*	1,5	2	2	2	1	1,5	1
	Trunkierung	1	1,5	2	0	2	1	0
	Speichern der Suchanfrage	2	2	2	2	2	1,5	0
	Inkrementelle Suche	0	0	2	1	0	0	2
	Durchsuchter Bereich	2	2	2	2	2	2	0
6.7	Erweiterte Suchmögl.	2	1	1	1	2	1,5	0
6.8	Retrievalmaße							
	Gesamttrefferzahl	1,5	2	0	1,5	1,2	1,5	0,5
	Precision	2	1,9	2	2	1,9	1,5	2
	Relativer Recall	2	1,7	2	2	1,7	0,5	2
	Σ	20,5	19,6	17,5	16	15,8	14	14

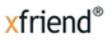
Tab. 23: Bewertung der Testergebnisse auf Suchmaschinen-Ebene

7.3 Desktop-Suchmaschinen-Ebene

								
	Kriterien	xfriend	Archivarius	Windows	Google	Copernic	FARR	Hulbee
6.9	Indexierung							
	Beeinflussungsmöglichkeiten	2	2	2	1,5	2	1,6	0
	Aktualität	2	2	2	1	2	1,5	0
	Indexierungsdauer	2	1,6	2	2	2	1,6	0
	Indexgröße	0	0,9	1,5	0,5	1	0,9	1,5
6.10	Vielseitigkeit							
	Art der Desktop-Suchmaschine	0,5	0	0,5	2	0,5	0,5	0,5
	Dateiformate*	2	1,5	1,5	1,5	1,5	1,5	1
	Spezielle Anwendungen	1,5	1,5	1,5	2	1,5	1	1,5
	Gelöschte und gepackte Dateien	2	2	0	1	0	0	0
6.11	Vertraulichkeit							
	Standpunkt Anbieter	1	0	0	0	0	0	0
	Schutz vor unbefugten Zugriffen	1,5	2	1	1	1	1,5	1,5
6.12	Darstellung der Suchergebnisse	2	1,5	2	1,5	2	1	1
	Σ	16,5	15	14	14	13,5	11,1	7

Tab. 24: Bewertung der Testergebnisse auf Desktop-Suchmaschinen-Ebene

7.4 Gesamtwertung

								
	Ebene	Copernic	xfriend	Windows	Google	Archivearius	FARR	Hulbee
7.1	Software-Ebene	11,5	6	10,5	11	6,5	10	7,5
7.2	Suchmaschinen-Ebene	19,6	20,5	17,5	16	15,8	14	14
7.3	Desktop-Suchmaschinen-Ebene	13,5	16,5	14	14	15	11,1	7
	Σ	44,6	43	42	41	37,3	34,1	28,5

Tab. 25: Zusammenfassung der Ergebnisse aller Ebenen

In der Gesamtbetrachtung der Ergebnisse des Softwarevergleichs wird deutlich, dass die Gesamtpunktzahlen der vier erstgenannten Suchmaschinen sehr eng beieinander liegen. Erst die letzten drei genannten Programme, Archivarius, Find and run robot und Hulbee weisen sowohl untereinander als auch im Vergleich zu Copernic, xfriend, Windows und Google größere Differenzen auf. Weiterhin ist zu erkennen, dass die einzelnen Desktop-Suchmaschinen im Hinblick auf die jeweiligen Untersuchungsebenen teilweise sehr unterschiedliche Resultate erzielen. Am deutlichsten wird dies am Beispiel der xfriend Desktop Search, welche trotz der schlechtesten Bewertung auf Software-Ebene insgesamt die zweithöchste Gesamtpunktzahl erreicht, was auf die sehr guten Ergebnisse im Suchmaschinen- und Desktop-Suchmaschinen-Bereich zurückzuführen ist.

8 Zusammenfassung und Ausblick

Ziel dieses Forschungsberichtes war es, die im Vorfeld ausgewählten Desktop-Suchmaschinen zu vergleichen. Um einen Vergleich durchführen zu können, muss zunächst eine Bewertungsgrundlage geschaffen werden, welche in der vorliegenden Arbeit durch den Kriterienkatalog dargestellt wird. Die gewählte Gliederung der Kriterien in die verschiedenen Bewertungsebenen Software-, Suchmaschinen- und Desktop-Suchmaschinen-Ebene bietet sich an, da so neben den für den Untersuchungsgegenstand, die (Desktop-)Suchmaschine, spezifischen Kriterien auch allgemeingültige Kriterien zur Bewertung der Softwarequalität aus Nutzersicht mit einbezogen werden.

Im Hinblick auf den eigentlichen Vergleich kann festgehalten werden, dass für einige der untersuchten Leistungsmerkmale durchaus größere Unterschiede zwischen den getesteten Suchmaschinen auftreten. In den meisten Fällen bestehen jedoch keine großen Differenzen, was deutlich wird, wenn man beispielsweise die Ergebnisse für den relativen Recall betrachtet. Auch andere Grundfunktionen, die von Suchmaschinen bekannt sind und demnach auch zu erwarten sind, wie die Eingrenzung der Suche nach bestimmten Gesichtspunkten, werden von nahezu allen Programmen gleichermaßen unterstützt. Auffällig ist auch, dass fast jede Suchmaschine ganz spezielle Vorteile hat, die keine andere bietet. So kann Copernic Videos direkt im Programm abspielen, bei Archivarius wird dafür jeder Suchbegriff der Ergebnisliste in einer anderen Farbe hervorgehoben, usw. Allgemein betrachtet fällt in der Testreihe weder ein eindeutiger Favorit auf, der jedes Kriterium unterstützt, noch kann eine Suchmaschine ausgemacht werden, die in jedem Punkt sehr schlechte Leistungen zeigt. Dieser Sachverhalt führt auch dazu, dass die Ergebnisse der Gesamtauswertung vor allem für die ersten vier Suchmaschinen Copernic, xfriend, Windows und Google sehr ähnlich ausfallen.

Im Gegensatz zur Forschung auf dem Gebiet der Suchmaschinen steht aktuell nur wenig wissenschaftliche Literatur für den spezielleren Bereich der Desktop-Suchmaschinen zur Verfügung. Vor allem fundierte Studien zum Vergleich verschiedener Anwendungen wurden bisher nicht durchgeführt. Die umfangreichste dokumentierte Evaluation, ist die in der vorliegenden Arbeit häufig als Referenz genannte Benchmark-Studie von Noda und Helwig aus dem Jahr 2005 [NoHe05]. Jedoch ist zu beachten, dass es sich hierbei in erster Linie um eine Dokumentation von Ergebnissen handelt. Zur Methodik werden keine An-

gaben gemacht. Gleiches gilt für die in PC-Zeitschriften veröffentlichten Testberichte, auf welche ebenfalls teilweise Bezug genommen wurde. Die vorliegende Arbeit unterscheidet sich von bisher veröffentlichten Untersuchungen durch die Aktualität der getesteten Programme und vor allem durch die Berücksichtigung der angewandten Methodik in der Dokumentation.

Gleichzeitig existieren kritische Stellen, für die, soweit möglich, noch Verbesserungsmöglichkeiten gefunden werden sollten. So ist es im Hinblick auf einige der untersuchten Kriterien sehr schwierig, eine eindeutige Bewertungsrichtung zu bestimmen. Am Beispiel der Gesamttrefferzahl wurde bereits im Text erklärt, dass sowohl eine niedrige, als auch eine hohe Gesamttrefferzahl als besseres Ergebnis angesehen werden können. Gleiches gilt für andere Kriterien, beispielsweise die Soundex-Suche. Kennt ein Benutzer die exakte Schreibweise des Suchbegriffs, so wird er eine Vielzahl von ähnlich geschriebenen Ergebnissen möglicherweise nicht als hilfreich, sondern als störend empfinden.

In Kapitel 6 der vorliegenden Arbeit wurde stets eine positive Bewertung gegeben, sobald die Suchmaschine ein Kriterium erfüllt. Oft treten jedoch „Mehrleistungen“ auf, beispielsweise das Highlighting in verschiedenen Farben, aber eine differenzierte Bewertung findet nicht statt. Die Schwierigkeit hierbei würde dann wiederum in der Wahl eines geeigneten Bewertungsmaßstabs bestehen.

Kritisch betrachtet werden muss auch das verwendete Bündelungssystem von Kriterien.

So werden einerseits für die Erfüllung eines einzelnen Kriteriums, dessen Prüfung meist mit relativ großem Arbeitsaufwand verbunden war (z.B. die Beachtung von Singular und Plural), zwei Punkte in der Gesamtauswertung vergeben. Andererseits wird die gesamte Tabelle 22 zur Darstellung der Suchergebnisse auf zwei Punkte normiert. Die Werte können zwar ohne großen Aufwand ermittelt werden, jedoch enthält diese Tabelle insgesamt sieben Einzelkriterien. Hier besteht die Gefahr einer Über- oder Untergewichtung einzelner Kriterien.

An eben erläuterten Punkt können weitere Forschungsthemen angesetzt werden. Für die Erstellung eines geeigneten Bewertungssystems würde sich eine weitere empirische Untersuchung anbieten. Durch die Befragung von Anwendern hinsichtlich für sie wichtiger und unwichtiger Funktionen einer Desktop-Suchmaschine könnten für jedes Kriterium Gewichtungsfaktoren festgelegt werden.

Betrachtet man die immer größer werdende Bedeutung von sozialen Netzwerken, sind Vermutungen durchaus berechtigt, dass diese auch im Hinblick auf Erweiterungen der Funktionalität von Desktop-Suchmaschinen neue, wesentliche Aspekte setzen können. Bereits heute ist ein Facebook-Gadget für die Seitenleiste von Google Desktop erhältlich, über welches das Eintreffen einer neuen Nachricht bei Facebook auf dem Desktop angezeigt wird. In Anbetracht dessen, werden natürlich die im Kriterienkatalog angeführten Bewertungskriterien in Zukunft dynamisch ergänzt werden müssen. Ebenso können Kriterien an Bedeutung verlieren und müssen somit wieder aus dem Katalog entfernt werden, was beispielsweise für die unter dem Begriff Effizienz zusammengefassten Kriterien zutreffen kann. Steigt die Leistungsfähigkeit von PCs in Zukunft stetig weiter, so wird z.B. die Bedeutung der Auslastung des Arbeitsspeichers durch das Programm oder der vom Programm benötigte Festplattenspeicher immer mehr an Relevanz für einen Vergleich verlieren und dafür andere Aspekte, wie bspw. die oben erwähnte Ankopplung an soziale Netzwerke in den Betrachtungsfokus rücken.

Literaturverzeichnis

- [BiCe09] Bischo-pinck, Y.; Ceyp, M.: Suchmaschinen-Marketing: Konzepte, Umsetzung und Controlling für SEO und SEM. Springer Verlag, 2. Auflage, Berlin, Heidelberg, 2009.
- [BlSc07] Blanc-Brude, T.; Scapin D. L.: What do People Recall about their Documents? Implications for Desktop Search Tools. 2007, <http://liris.cnrs.fr/~yprie/ens/07-08/Code-traces/PDF/p102-blanc-brude.pdf>, Abruf am 2009-11-04.
- [BuSc06] Burmeister, O.; Schirrmann, A.: Desktop Search Engines. Kaiserslautern, 2006, http://www.informatik.fh-kl.de/~amueller/Desktop%20SearchEngines_Schirrmann_Burmeister.pdf, Abruf am 2010-06-20.
- [CNSe04] Fazli Can, Rabia Nuray, Ayisigi B. Sevdik: Automatic performance evaluation of Web search engines. In: Information Processing & Management. 2004, Nr. 3, S. 495-514.
- [Cole05] Cole, B.: Search engines tackle the desktop. In: Computer, Bd. 38, Nr. 3, 2005, S. 14-17.
- [Depa09] De, P.: FARR Is a Launchy Alternative for Power Users. 2009, <http://www.pallab.net/2009/12/21/farr-is-a-launchy-alternative-for-power-users/>, Abruf am 2010-06-08.
- [Endr03] Endres, A.: Softwarequalität aus Nutzersicht und ihre wirtschaftliche Bewertung. In: Informatik-Spektrum. Nr. 26, 2003, S. 20-25.
- [Folm05] Folmer, E.: Software Architecture – Analysis of Usability. Enschede, 2005, <http://dissertations.ub.rug.nl/FILES/faculties/science/2005/e.folmer/thesis.pdf>, Abruf am 2010-10-22.
- [GrFr04] Grossman, D.; Frieder, O.: Information Retrieval - Algorithms and Heuristics. Springer Verlag, 2. Auflage, Dordrecht 2004.

- [KiJu01] Kirchner, L; Jung, J.: Ein Bezugsrahmen zur Evaluierung von UML-Modellierungswerkzeugen. Koblenz, 2001, http://www.wi-inf.uni-duisburg-essen.de/FGFrank/documents/Arbeitsberichte_Koblenz/Nr26.pdf. Abruf am 2009-11-04.
- [Koch07] Koch, D.: Suchmaschinen-Optimierung: Website-Marketing für Entwickler. Addison-Wesley Verlag, München, 2007
- [Lewa07] Lewandowski, D.: Mit welchen Kennzahlen lässt sich die Qualität von Suchmaschinen messen? Köln, 2007, http://www.bui.haw-hamburg.de/fileadmin/user_upload/lewandowski/doc/Kennzahlen_Qualitaet_von_Suchmaschinen.pdf, Abruf am 2010-06-22.
- [LeHö07] Lewandowski, D.; Höchstötter, N.: Qualitätsmessung bei Suchmaschinen – System- und nutzerbezogene Evaluationsmaße. In: Informatik-Spektrum, Bd. 30, Nr. 3, 2007, S. 159-169
- [LSSW] Lu, C.; Shukla, M.; Subramanya, S. H.; Wu, Y: Performance Evaluation of Desktop Search Engines. In: IEEE International Conference on Information Reuse and Integration (IRI 2007), http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4296606, Las Vegas, 2007, Abruf am 2010-06-25.
- [Mark10] Markscheffel, B.: Information Retrieval – Inhaltserschließung. 2010, <http://informationsmanagement.wirtschaft.tu-ilmelmenau.de/lehreundforschung/veranstaltung-ss10/documents/IR2.pdf>.
- [Nefz09] Nefzger, W.: Vergleichstest Desktop-Suchprogramme – Auf dem Index – Copernic Professional 3.1. 2009, <http://software.magnus.de/tools/artikel/vergleichstest-desktop-suchprogramme.2.html>, Abruf am 2010-06-08
- [NoHe05] Noda, T.; Helwig, S.: Benchmark Study of Desktop Search Tools – There`s more to search than Google & Yahoo! - An Evaluation of 12 Leading Desktop Search Tools. Wisconsin, 2005, http://www.uwebi.org/reports/desktop_search.pdf. Abruf am 2009-10-28.
- [Padd09] Paddock, B.: Start++ for Windows 7 test build 2. 2009, <http://brandontools.com/files/folders/sppdevbuilds/entry743.aspx>, Abruf am 2010-06-22.

- [PiWo09] Pieper, D.; Wolf, S.: Wissenschaftliche Dokumente in Suchmaschinen. In: Lewandowski, D. (Hrsg.): Handbuch Internet-Suchmaschinen - Nutzerorientierung in Wissenschaft und Praxis. AKA Verlag, Heidelberg, 2009, S. 356-374.
- [Rems09] Remse, C.: Die 10 besten Desktop-Such-Tools. 2009, http://www.pcwelt.de/start/software_os/windows/news/193713/die_10_besten_desktop_such_tools/. o.O. 2009, Abruf am 2010-10-28.
- [Schü08] Schütz, A.: Die Gefahren des Google-Imperiums. 2008, http://www.silicon.de/software/business/0,39039006,39161360,00/die+gefahren+des+google_imperiums.htm, Abruf am 2010-06-25.
- [Stoc07] Stock, W.G.: Information Retrieval – Informationen suchen und finden. Oldenbourg, München, Wien, 2007. [Sünk09] Sünkler, S.: Usability-Evaluation von Suchmaschinen. GRIN Verlag, Norderstedt, 2009.
- [THNi04] Toutenburg, Heumann, Nittner: Statistische Methoden bei unvollständigen Daten. München, 2004, http://epub.ub.uni-muenchen.de/1750/1/paper_380.pdf, Abruf: 2010-06-21.
- [WeWa02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. 2002, <http://www.misq.org/archivist/vol/no26/issue2/GuestEd.pdf>, Abruf am 2009-11-30.
- [Woll05] Wolling, J.: Suchmaschinen? - Selektiermaschinen! 2005, http://www.wolling.de/pdfs/wolling_suchmaschinen_selektiermaschinen.pdf, Abruf am 2009-11-04.

Anhang

Anhang A: Ergänzende Bemerkungen zur Testkollektion

- Speicherort der Testkollektion: C:\Benutzer\Benutzername

- Die Dateien der Testkollektion werden auf Unterverzeichnisse des genannten Speicherortes in unterschiedlicher Gliederungstiefe verteilt. Nicht zur Testkollektion zählt das Unterverzeichnis „AppData“, welches unter C:\Benutzer\Benutzername\AppData zu finden ist.

- Kann der Nutzer auswählen, welche Verzeichnisse von der Desktop-Suchmaschine indiziert werden sollen, wird stets das komplette Laufwerk C: ausgewählt.

- Alle weiteren standardmäßig auf dem Rechner verfügbaren Ordner und deren Inhalt, wie beispielsweise C:\Windows, werden so auch im Test berücksichtigt. Da keine Veränderungen an der Anfangseinstellung des Testsystems vorgenommen werden, mit Ausnahme der Installation der in Tabelle 2-2 genannten Anwendungssoftware, findet an dieser Stelle keine detailliertere Auflistung des Inhalts statt.

- Das Testsystem beinhaltet keine externen Laufwerke.

Anhang B: Benutzeroberflächen der getesteten Suchmaschinen

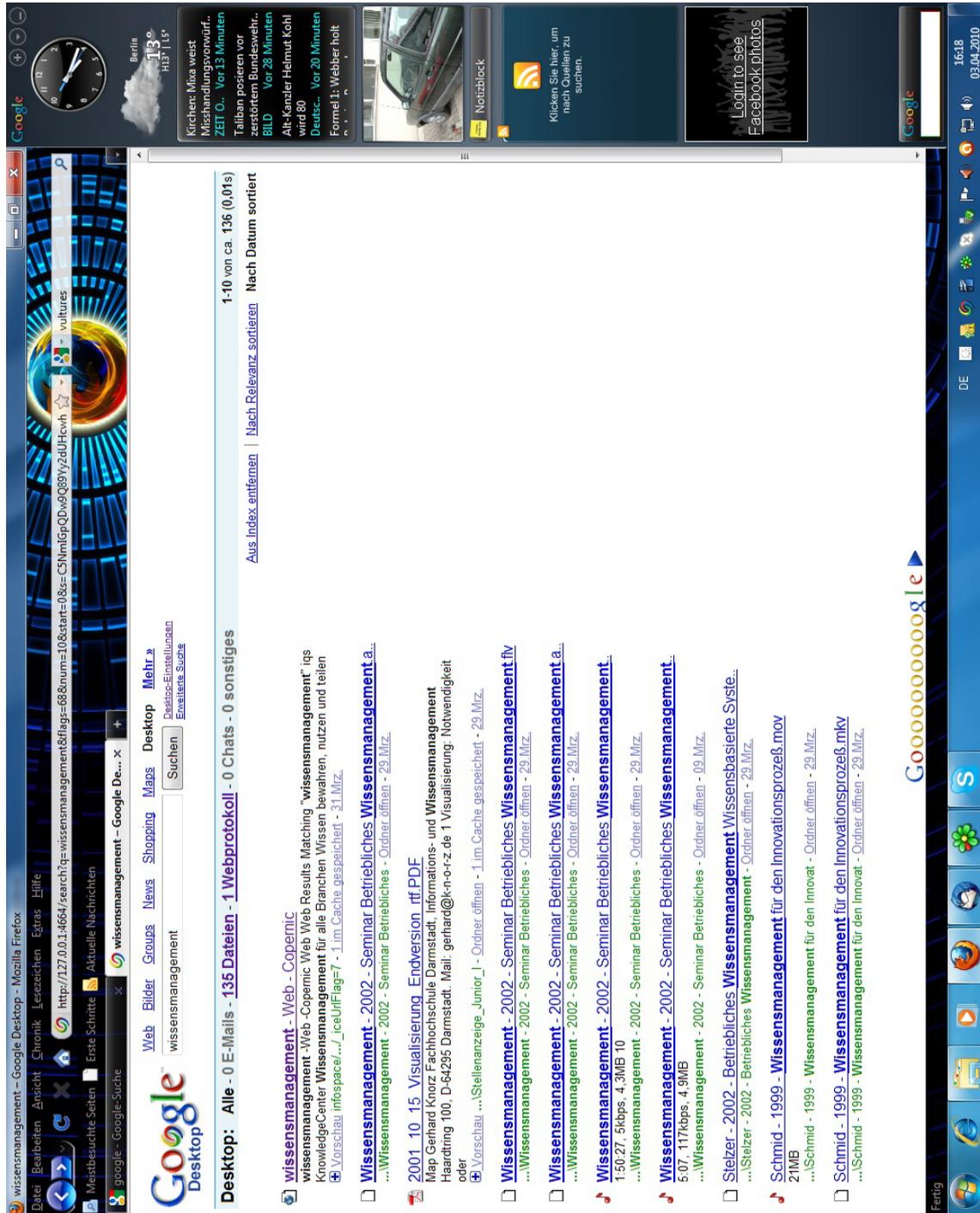
Windows 7-Suche

Darstellung: Suchergebnisseite mit Highlighting des Suchbegriffs.



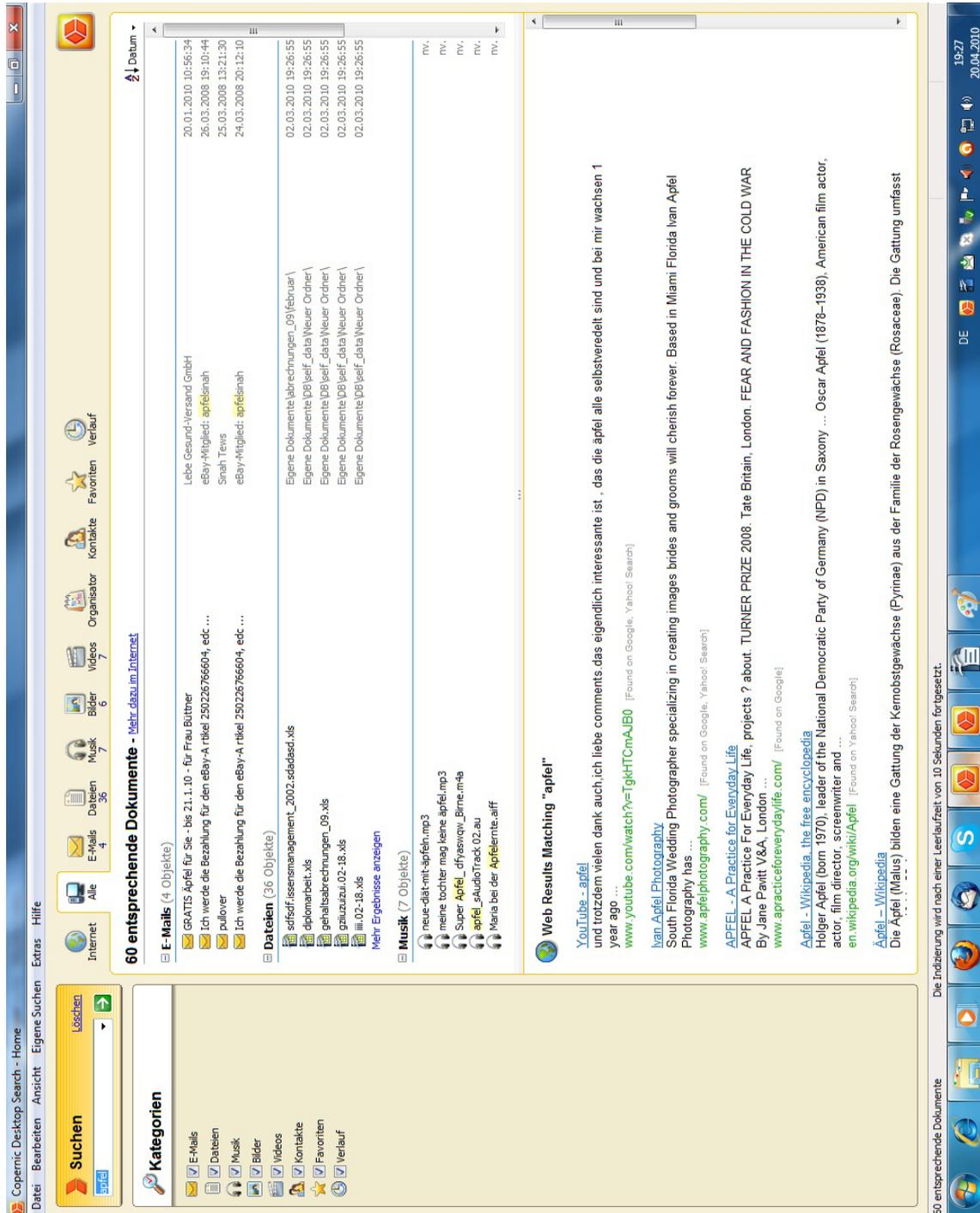
Google Desktop

Darstellung: Suchergebnisseite und Seitenleiste mit Standardeinstellungen.



Copernic Desktop Search

Darstellung: Suchergebnisseite mit ausgewähltem Menüpunkt „Alle anzeigen“.



xfriend

Darstellung: Suchergebnisseite mit hervorgehobenem Suchbegriff und aufgeklapptem Dropdown-Menü zur Ergebnisfilterung.



Archivarius 3000

Darstellung: Suchergebnisseite mit hervorgehobenem Suchbegriff und Einblendung der Vorschau im rechten Fenster.

The screenshot displays the Archivarius 3000 search results page. The search term is 'wissensmanagement'. The results list shows various PDF files related to 'Wissensmanagement' with a 'Bewertung' (rating) column. A preview window on the right shows the content of the selected file, highlighting the search term 'Wissensmanagement' in yellow. The preview text includes: 'Betriebliches Wissensmanagement / wissensbasierte Systeme (Überblick über Techniken)', 'Prof. Dr. D. Stelzer', 'Betriebliches Wissensmanagement / wissensbasierte Systeme Überblick über Techniken des Wissensmanagements', 'Prof. Dr. Dirk Stelzer', 'Technische Universität Ilmenau Institut für Wirtschaftsinformatik Fachgebiet Informationsmanagement', 'TUI, FG IM, WS 2001/2002', 'Lernziele', and 'sie haben einen Überblick über verschiedene Techniken des Wissensmanagements. Sie können Anforderungen an Techniken zur Unterstützung des betrieblichen Wissensmanagements formulieren. Sie können erörtern, welchen Techniken in welchem Maße helfen können, welche (Teil-)Ziele des Wissensmanagements zu erreichen.'

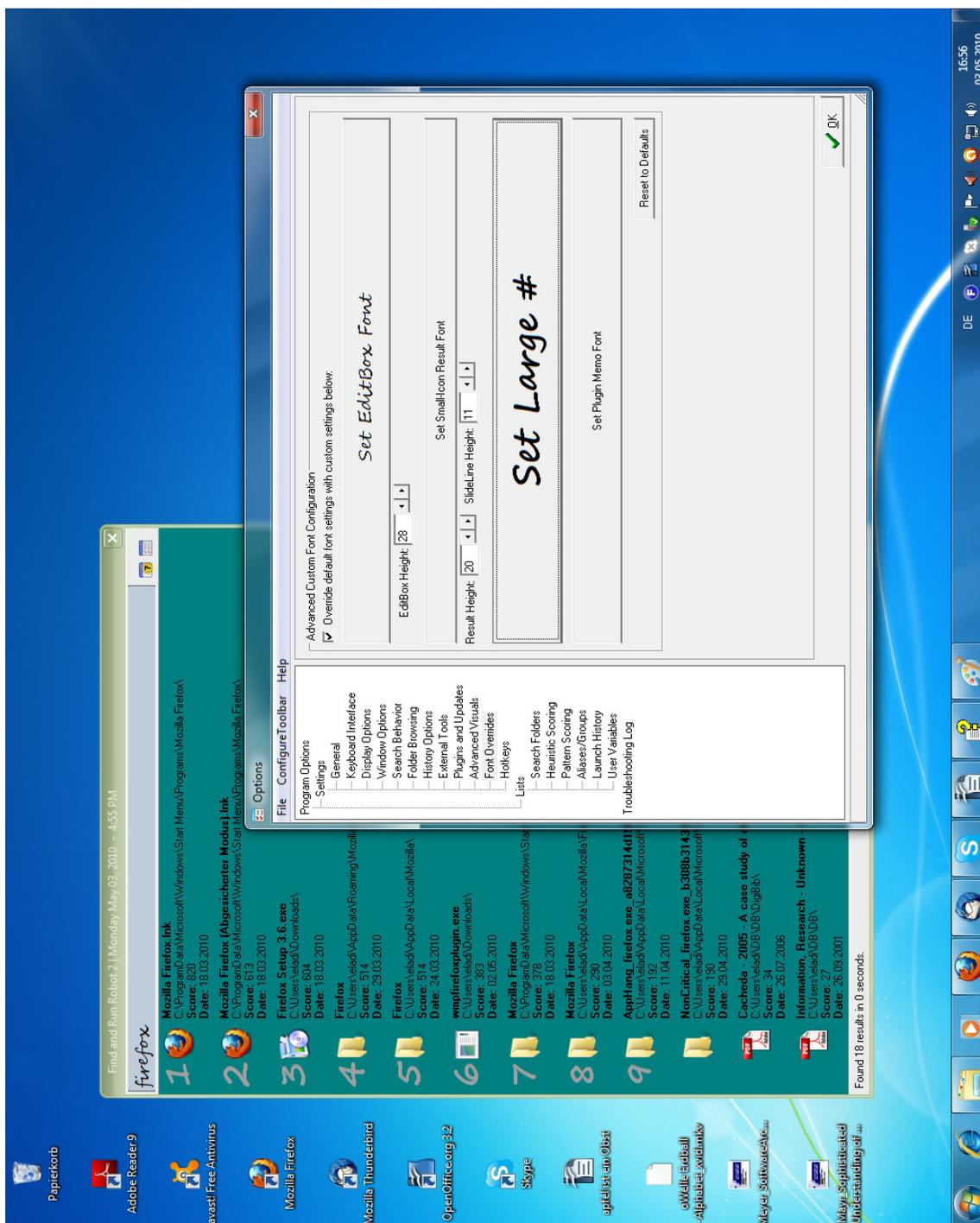
Hulbee Desktop

Darstellung: Suchergebnisseite mit standardmäßig angezeigter Data-Cloud links und Filtermöglichkeiten nach Sprache und Typ rechts.

The screenshot displays the Hulbee Desktop search engine interface. At the top, there is a search bar with the text "Suchbegriff (e)" and a search button. Below the search bar, the results are listed as "Ergebnisse für **information** gefunden: 1858". The main content area shows a list of search results, each with a document icon, a title, and a brief description. The results are filtered by language and document type. On the right side, there are two panels: "Anzeigen nach Sprache:" with language options (All, Deutsch, Invariante Sprache, Nederlands, Magyar, Português, Français, Español, English) and "Dokumente nach Typ" with document type options (All, Html, Pdf, Jpeg, Jst, Doc, Jtm, Jif, Docx). The bottom of the interface features a navigation bar with various icons and a footer with the text "Ein Produkt der Grossbay AG. Alle Rechte vorbehalten. ©2009 Grossbay AG. Besuchen Sie www.hulbee.com | www.grossbay.com für mehr Informationen."

Find and run robot

Darstellung: Suchergebnisseite und „Optionen“-Fenster; gerade durchgeführt wird die Änderung der Schriftart des Suchbegriffs; Hintergrundfarben wurden von grau/weiß bereits auf die angezeigten Grün- und Blautöne geändert;



Anhang C: Erläuterungen zur Standard-Tabellenform und Webadressen der Suchmaschinen-Anbieter

Symbolik der Bewertungstabellen

+	Positive Bewertung; das jeweilige Kriterium wird unterstützt; 1 Punkt; *
-	Negative Bewertung; das jeweilige Kriterium wird nicht unterstützt; 0 Punkte; *
+/-	Das Kriterium wird nur teilweise unterstützt; nähere Erläuterungen sind jeweils im Text zu finden; 0,5 Punkte;
Σ	Gesamtsumme der für die Einzelkriterien vergebenen Bewertungspunkte pro Suchmaschine;
%	Für das jeweilige Kriterium kann für die entsprechende Suchmaschine keine Bewertung vorgenommen werden;

* Die Bewertung erfolgt hierbei stets orientiert am gewünschten Ergebnis. Werden beispielsweise die Einschränkungen der Testversion geprüft (5.1), so ist gewünscht, dass keine Einschränkungen vorliegen. Alle Suchmaschinen, für die keine Einschränkungen vorliegen, werden folglich positiv mit „+“ bewertet. Liegen hingegen welche vor, dann erfolgt eine Bewertung mit „-“.

Webadressen der Suchmaschinen-Anbieter

Google	http://desktop.google.com
Copernic	http://www.copernic.com/en/products/desktop-search
xfriend	http://www.xfriend.de/de/privat/uebersicht/
Archivarius	http://www.likasoft.com/de/document-search/
Hulbee	http://company.hulbee.com/products_and_solutions/hulbee_desktop/
FARR	http://www.donationcoder.com/Software/Mouser/findrun/

Anhang D: Grundlagen zur Messung der Effizienz und Messergebnisse

Exakte Messwerte zur Größe von Installationsdatei und Programmordner

	Windows	Google	Copernic	xfriend	Archivar.	Hulbee	FARR
Installations-datei	%	1,9MB	8,0MB	58,0MB	4,1MB	27,9MB	5,9MB
Programm-ordner	%	8,23MB	32,6MB	163,0MB	10,1MB	56,7MB	10,4MB

Grundlagen der Messung:

- Für die Größe der Installationsdatei: Unter der Bezeichnung „Größe“ angegebene Wert in den Eigenschaften der heruntergeladenen Installationsdatei
- Für die Größe des Programmordners: Unter der Bezeichnung „Größe“ angegebener Wert in den Eigenschaften des Ordners „*Programmname*“ unter dem Pfad C:/Programme x86/Programmname

Exakte Messwerte zur Speicherauslastung

	Windows	Google	Copernic	xfriend	Archivarius	Hulbee	FARR
CPU P	4,75%	5,73%	10,43%	9,9%	16,47%	9,63%	0,5%
RAM P	1,39 GB	1,63 GB	1,25 GB	1,60 GB	1,50 GB	1,52 GB	1,32 GB
CPU S	25%	15%	16%	30%	32%	42%	51%
RAM S	1,40 GB	1,64 GB	1,29 GB	1,58 GB	1,51 GB	1,54 GB	1,31 GB
CPU I	49%, 0%	100%, 37%	79% 8% 25% 0%*	79%, 1%	76%, 18%	61%, 6%	%
RAM I	1,65 GB	1,53 GB	1,36 GB	1,66 GB	1,63 GB	1,45 GB	%

Details zur Messung:

- Hilfsinstrument: Windows Task Manager, Option „Leistung“.
- Ausgangsbedingungen/geöffnete Programme: zu testende Desktop-Suchmaschine, Mozilla Firefox mit einem Tab, Mozilla Thunderbird, ein Microsoft Word-Dokument.
- *Programmbetrieb*
 - Beobachtungszeitraum: 10min

- RAM-Auslastung: Die Summe aller auftretenden Werte wird dividiert durch die Anzahl der gemessenen Werte.
Mehrmals vorkommende Werte werden nur einmal beachtet.
 - CPU-Auslastung: s. RAM-Auslastung
- *Suchvorgang:*
- Beobachtungszeitraum: unmittelbar nach Bestätigung der Suchanfrage
 - RAM-Auslastung: Messung des angezeigten Wertes unmittelbar nach Bestätigung der Suchanfrage mit „Enter“.
 - CPU-Auslastung: s. RAM-Auslastung
- *Indexierung:*
- Beobachtungszeitraum: erste 15min der Indexierung, bzw. Gesamtdauer wenn Indexierungsdauer < 15min (bei RAM); erste Minute der Indexierung (bei CPU)
 - RAM-Auslastung: s. Programmbetrieb
 - CPU-Auslastung: Messung des kleinsten und größten auftretenden Wertes. Für die Grafik wird aus beiden der Mittelwert gebildet
 - Für Copernic wurden hier vier Werte gemessen. Grund ist die im Hauptteil erläuterte Verwendung zwei verschiedener Indexierungsgeschwindigkeiten (geringer Ressourcenbedarf: 0% und 8%; Verwendung aller Ressourcen: 25% und 79%).
 - Für die Grafik wurde der Mittelwert für die ressourcensparende Variante berechnet, da diese voreingestellt ist und immer verwendet wird, wenn der Benutzer keine selbsttätigen Änderungen vornimmt

Anhang E : Grundlagen zur Messung der Indexgröße

Für die Größe des Index wird der unter der Bezeichnung „Größe“ angegebene Wert in den Eigenschaften der in Fettdruck hervorgehobenen Verzeichnisse angegeben:

Archivarius	C:\Benutzer\Benutzername\AppData\Roaming\Archivarius3000
Windows	C:\Program Data\Windows\Search\Windows
Google	C:\Benutzer\Benutzername\AppData\local\Google\GoogleDesktop
Copernic	C:\Benutzer\Benutzername\AppData\local\Copernic\DesktopSearch\Index
Hulbee	C:\Benutzer\Benutzername\AppData\Roaming\Grossbay
xfriend	C:\Benutzer\Benutzername\AppData\Roaming\.\dot\xfriend\index

Der Ordner „Index“ ist nicht für jede Suchmaschine verfügbar. Aus Gründen der Vergleichbarkeit ist jedoch festzuhalten, dass für Copernic und xfriend kein nennenswerter Größenunterschied zwischen „Index“ und dem übergeordneten Verzeichnis „Desktop Search“ bzw. „xfriend“ besteht.

Anhang F: Auszug aus der Liste von xfriend unterstützter Dateiformate

xfriend 2.9 - Unterstützte Dateiformate
 Stand: 30.03.2010

Suchen

Anwendung	Version	Suchen	Vorschau	Thumbnail	Endung	Metadaten
standardmäßig indexiert						
MS Word	95-2007	++	++	++	doc, docx, docm	Titel, Thema, Autor, Stichwörter, Kommentare
MS Word Vorlage	95-2007	++	++	++	dot, dotx, dotm	Titel, Thema, Autor, Stichwörter, Kommentare
MS Excel	4.0, 95-2007	++	++	++	xls, xlsx, xism	Titel, Thema, Autor, Stichwörter, Kommentare
MS Excel Vorlage	4.0, 95-2007	++	++	++	xlt, xltx, xltn	Titel, Thema, Autor, Stichwörter, Kommentare
MS Powerpoint	95-2007	++	++	++	ppt, pptx, pptm	Titel, Thema, Autor, Stichwörter, Kommentare
MS Powerpoint Vorlage	95-2007	++	++	++	pot, potx, potm	Titel, Thema, Autor, Stichwörter, Kommentare
MS Powerpoint Bildschirmpräsentation	95-2007	++	++	++	pps, ppsx, ppsm	Titel, Thema, Autor, Stichwörter, Kommentare
MS Access	bis 2.0	++	++	++	mdb	Titel, Thema, Autor, Stichwörter, Kommentare
MS Outlook	ab Version 95	++	++	++	pst (über Outlook), eml, msg	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Writer	1.x	++	++	++	sww	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Writer Vorlage	1.x	++	++	++	stw	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Writer Global	1.x	++	++	++	svg	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Calc	1.x	++	++	++	sxc	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Calc Vorlage	1.x	++	++	++	s1c	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Impress	1.x	++	++	++	sxl	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Impress Vorlage	1.x	++	++	++	s1l	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Draw	1.x	++	++	++	svd	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Draw Vorlage	1.x	++	++	++	svd	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Writer	2.0 - 3.x	++	++	++	odt	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Writer Vorlage	2.0 - 3.x	++	++	++	odt	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Writer Global	2.0 - 3.x	++	++	++	odm	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Calc	2.0 - 3.x	++	++	++	ods	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Calc Vorlage	2.0 - 3.x	++	++	++	ods	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Impress	2.0 - 3.x	++	++	++	odp	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Impress Vorlage	2.0 - 3.x	++	++	++	odp	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Draw	2.0 - 3.x	++	++	++	odg	Titel, Thema, Autor, Stichwörter, Kommentare
OpenOffice Draw Vorlage	2.0 - 3.x	++	++	++	odg	Titel, Thema, Autor, Stichwörter, Kommentare
Text		++	++	++	txt	Titel, Thema, Autor, Stichwörter, Kommentare
RTF		++	++	++	rtf	Titel, Thema, Autor, Stichwörter, Anwendung, Ersteller
PDF		++	++	++	pdf	Titel, Thema, Autor, Stichwörter, Anwendung, Ersteller
Shell Skript		++	++	++	sh	Titel, Thema, Autor, Stichwörter, Anwendung, Ersteller
HTML		++	++	++	html, htm	Titel, Thema, Autor, Stichwörter, Anwendung, Ersteller
Comma Separated Values (CSV)		++	++	++	csv	Titel, Thema, Autor, Stichwörter, Anwendung, Ersteller
XML Dokumente		++	++	++	xml	Titel, Thema, Autor, Stichwörter, Anwendung, Ersteller
Postscript (PS)		++	++	++	ps	Titel, Thema, Autor, Stichwörter, Anwendung, Ersteller
MP3		++	++	++	mp3	Titel, Thema, Autor, Stichwörter, Anwendung, Ersteller
OGG		++	++	++	ogg	Titel, Thema, Autor, Stichwörter, Anwendung, Ersteller

©2010 xdot GmbH | Felstiege 78 | 48161 Münster
 Fon: +49 1805 4389368 - 100 | Fax: +49 1805 4389368 - 200 | Internet: www.xdot.de | Email: info@xdot.de
 xfriend® ist eingetragenes Markenzeichen der xdot GmbH.

Anhang G: Grundlagen zur Messung der unterstützten Dateiformate

Zu jedem in der Testkollektion vertretenen Dateiformat wird eine passende Suchanfrage gestellt. Hierbei handelt es sich um einen Dateinamen, der in Zusammenhang mit diesem Format auftritt. Um beispielsweise die Unterstützung des JPEG-Formats zu prüfen wird das in der Testkollektion vorhandene „asien.jpg“ über die Anfrage „Asien“ gesucht. Für jedes Format wurde vor Beginn der Testphase ein bestimmter Suchbegriff festgelegt.

Unterstützt eine Suchmaschine auch die Möglichkeit direkt nach den Dateitypen zu suchen, also z.B. die Ermittlung von JPEG-Dateien über die Anfrage „jpg“, so kann vereinfachend diese Option verwendet werden.

Anhang H: Normierung der Ergebnisse für die Auswertung

Normierung auf das Maximum:

Sie wird durchgeführt, wenn der gewünschte Wert der größtmögliche Wert ist.

Erläutert wird der Sachverhalt nachfolgend am Beispiel des Kriteriums „Unterstützung der Operatoren AND, OR, NOT NEAR“ (Kapitel 5.6):

Eine Suchmaschine kann hier, wenn sie alle genannten Operatoren unterstützt, maximal mit vier „+“-Symbolen, also mit vier Gesamtpunkten bewertet werden. Dieses Maximum von 100% entspricht in der Auswertungstabelle dem größtmöglichen Wert von zwei Punkten. Aus der nachfolgenden Tabelle erkennt man, dass auch eine Suchmaschine, die im fünften Kapitel mit 3,5 Punkten bewertet wurde in der Auswertung zwei Punkte erhält

Werte in Kapitel 5	Prozentsatz	Bewertung in Kapitel 6
3,21 - 4 Punkte	100% - 80%	2
2,41 - 3,2 Punkte	79% - 60%	1,5
1,61 - 2,4 Punkte	59% - 40%	1
0,81 - 1,6 Punkte	39% - 20%	0,5
0 - 0,8 Punkte	19% - 0%	0

Die mit „Prozentsatz“ und „Bewertung in Kapitel 6“ überschriebenen Spalten werden für jedes Kriterium identisch verwendet, wohingegen die „Werte in Kapitel 5“ stets neu berechnet werden.

Normierung auf das Minimum:

Sie wird durchgeführt, wenn der gewünschte Wert der niedrigste Wert ist, z.B. bei der Speicherauslastung (Kapitel 5.2). Der niedrigste gemessene Wert entspricht hier somit 100% bzw. der Obergrenze zum Erreichen der zwei Punkte.

Anhang I: Erläuterung abweichender Bewertungen

Usability:

→ Windows: 1,5 Punkte statt 2 Punkte

Grund: versteckt liegende Funktionen

→ FARR: 1 Punkt statt 2 Punkte

Grund: sehr minimalistisches Design und weniger Funktionen als die übrigen Suchmaschinen; außerdem schwere Verständlichkeit der englischsprachigen Programmoptionen;

→ Archivarius: 1 Punkt statt 2 Punkte

Grund: Orientierung im Programm ist schwierig; minimalistischer Farbeinsatz; überfordernde Funktionsanzahl

Standardverknüpfung

- AND: 2 Punkte, da von Google-Websuche verwendet und somit erwartete Verknüpfung
- Nicht eindeutig: 1, 5 Punkte; nicht eindeutig als AND und OR identifizierbar
- OR: 1 Punkt, da von den Nutzern weniger erwartet als AND;

Dateiformate

- Unterstützung der in der Testkollektion enthaltenen Formate: 1,5 Punkte
- eines oder mehrere Formate werden nicht unterstützt: 1 Punkt
- Bonus von 0,5 Punkte für ausführliche Darstellung unterstützter Formate in PDF-Datei: 2 Punkte