Fachhochschule Fulda
University of Applied Science

Fachbereich Angewandte Informatik

Diplomarbeit

Entwurf und Modellierung der Personalisierung einer Relevanzbewertung für Dokumente – Theorie und praktische Implementation

Bearbeiter: Andreas Lennartz (164580)
Referent: Hans Helmut Paul
Korreferent: Valerij Harlamow

Fulda, den 25.7.05

Sperrvermerk

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

1.1 Problemstellung

Die Anzahl an verfügbaren Informationen im Internet steigt stetig an. Ein Durchsuchen dieser Informationen ist ohne geeignete Suchdienste kaum möglich. Doch mit der zunehmenden Menge an Informationen haben inzwischen auch die bestehenden Suchdienste immer öfter Probleme, die für den Informationssuchenden relevanten Informationen zu finden und von den nicht relevanten Informationen zu unterscheiden.

Dabei werden Anfragen an Suchdienste und deren Suchmaschinen in der Regel durch kurze und prägnante Terme formuliert. Die Suchmaschine versucht dann, anhand dieser Terme die für den Nutzer des Suchdienstes wahrscheinlich relevantesten Informationen zu finden. Informationen werden in der Regel durch Dokumente (wie z.B. Webseiten) dargestellt. Meist wird eine recht große Anzahl an Dokumenten gefunden, die dann nach ihrer Relevanz für den Nutzer absteigend sortiert und angezeigt werden. Die persönlichen Vorlieben und Interessen des Suchenden werden in der Regel nicht berücksichtigt.

Da bestehende Suchmaschinen keine weiteren Informationen über den Nutzer selbst einholen, kann die Relevanz der Dokumente nicht in geeigneter Weise individuell für einen Nutzer ermittelt werden. So kann die Sortierung nach der Relevanz für den Nutzer ohne weitere Informationen über diesen nicht optimal durchgeführt werden. Das folgende Beispiel verdeutlicht dies: Die Abfrage mit dem Suchbegriff “Apple” eines englischsprachigen Nutzers kann sich auf Dokumente beziehen, die sich mit der Frucht befassen, während ein anderer Nutzer Dokumente sucht, die sich mit dem gleichnamigen Computer-Hersteller thematisch auseinandersetzen. Eine normale Suchmaschine liefert für beide Nutzer die gleichen Ergebnisse.

„As the amount of information on the Web increases rapidly, it creates many new challenges for Web search. When the same query is submitted by different users, a typical search engine returns the same result, regardless of who submitted the query. This may not be suitable for users with different information needs. “[1,S.2]

Die Möglichkeit, die persönlichen Vorlieben und Interessen des Suchenden bei einer Suchabfrage zu berücksichtigen, erscheint somit sinnvoll. So könnte beim Auffinden der relevanten Dokumente dies mit einbezogen und Dokumente individuell nach Ihrer Relevanz für den jeweiligen Nutzer bewertet werden.

Um dieses Problem zu lösen, existieren bereits benutzerverhaltensorientierte Ansätze und Lösungen im Bereich des Information Retrieval, die eine so genannte „Personalisierung" von Dokumenten vornehmen. Diese haben bis jetzt noch nicht die nötige Akzeptanz beim Benutzer gefunden. Häufige Probleme beim Erstellen solcher benutzerverhaltensorientierten Lösungen sind zum einem das Erstellen eines „Profils" eines Nutzers, mit dem seine Vorlieben und Interessen beschrieben werden, sowie die Bewertung der Relevanz eines Dokumentes mit Hilfe dieses Profils.

Die Firma Global Brain Network GmbH mit Sitz in Fulda ist seit über fünf Jahren mit der Entwicklung einer benutzerverhaltensorientierten Suchmaschine beschäftigt. Der Name dieser Suchmaschine, welche bereits öffentlich zugänglich ist, lautet „Fooxx"[1]. Ein Nutzer kann sich bei Fooxx registrieren, sein individuelles Profil erstellen und zunächst speichern. Der Suchdienst Fooxx liefert bei einer Suchanfrage eines registrierten Nutzers eine Ergebnisliste, die auf Basis des angegebenen Profils personalisiert wurde. Leider besitzt das bestehende Fooxx bei der Erstellung und Verarbeitung des Profils sowie bei der Personalisierung der Dokumente einige Schwachstellen.

In der vorliegenden Arbeit werden diese Schwachstellen aufgezeigt und behoben. Dazu wird ein Algorithmus entwickelt, welcher die Bewertung der Relevanz eines Dokumentes auf der Grundlage des Profils des suchenden Nutzers vornimmt. Weiterhin wird beschrieben, wie dieses Profil ermittelt sowie eine Personalisierung mit Hilfe des entwickelten Algorithmus durchgeführt wird. Zudem wird eine Testumgebung entwickelt, mit welcher der entwickelte Algorithmus und ggf. auch andere Algorithmen auf ihre Funktionsweise hin überprüft werden können. Es werden Tests durchgeführt, die untersuchen, ob der entwickelte Algorithmus sein gewünschtes Ergebnis, nämlich die personalisierte Bewertung der Relevanz von Dokumenten, erzielt. Der entwickelte Relevanzbewertungsalgorithmus soll bei

[1] http://www.fooxx.com

erfolgreichen Tests in das bestehende System von Fooxx eingebaut werden und die bestehende Profilerstellung und Personalisierung ersetzen.

Ziel dieser Arbeit ist es somit, einen funktionsfähigen benutzerverhaltensorientierten Relevanzbewertungs-Algorithmus zu entwerfen, zu implementieren, zu testen und ggf. zu optimieren. Sollte der Algorithmus die gewünschten Anforderungen erfüllen, ist eine Integration in das bestehende Fooxx geplant.

1.2 Aufbau der Arbeit

Im Anschluss an die Einleitung werden in *Kapitel 2* die Grundlagen des Information Retrievals erläutert. Gängige Modelle von IR-System, die sich mit dem Auffinden von relevanten Dokumenten in einer Dokumentenmenge beschäftigen, werden vorgestellt. Ein Überblick über bekannte Suchdienste im Internet wird gegeben. Neben den Konzepten der Personalisierung von Dokumenten wird das System der sozialen Filterung, auf welches der zu entwickelnde Algorithmus basiert, detaillierter erläutert.

In *Kapitel 3* wird der Suchdienst „Fooxx“ vorgestellt und näher erläutert. „Fooxx“ ist eine benutzerverhaltensorientierte Suchmaschine, welche Dokumente anhand den Vorlieben und Interessen des Suchenden sortiert. Es wird erläutert, wie die Erstellung eines Profils eines Nutzers abläuft, wie weitere Daten über den Nutzer und seinen aufgerufenen Dokumenten gesammelt und wie diese gesammelten Informationen zur Personalisierung der Suchergebnisse verwendet werden. Die Schwächen des bisherigen Systems der Profilerstellung und Personalisierung werden konkretisiert.

In *Kapitel 4* wird ein Algorithmus entworfen, der eine Relevanzbewertung von Dokumenten aufgrund des Profils eines Nutzers umsetzt. Dies wird durch die Einbeziehung der Dokumente anderer Nutzer umgesetzt, die das zu bewertende Dokument ebenfalls genutzt haben. Ziel dieser Relevanzbewertung ist, damit die im vorherigen Kapitel aufgezeigten Schwachstellen des bestehenden Fooxx zu beheben.

In *Kapitel 5* wird dieser Algorithmus praktisch implementiert. Weiterhin wird eine Testumgebung modelliert und umgesetzt, um die Auswirkungen des Algorithmus näher zu untersuchen. Dabei werden die Daten des bestehenden Fooxx als Grundlage genommen. Eine Anbindung an verschiedene Suchmaschinen erfolgt, um eine Vorauswahl von sortierten Dokumenten zu erhalten.

In *Kapitel 6* wird der Algorithmus in verschiedenen Testreihen auf seinen Zweck untersucht. Es wird analysiert, ob eine Personalisierung von Suchergebnissen feststellbar ist und inwieweit relevante Seiten in der Bewertung berücksichtigt werden.

In *Kapitel 7* werden die Abfragezeiten und mögliche sowie durchgeführte Optimierungen erläutert.

In *Kapitel 8* wird ein abschließendes Resümee über die gebrachte Arbeit gegeben und es werden Vorschläge für Verbesserungen gemacht. Es wird diskutiert, ob und wie der entwickelte Algorithmus in die bestehende Lösung von Fooxx integriert werden kann.

2 Grundlagen des Information Retrieval

Dieses Kapitel erläutert die grundlegenden Prinzipien des Information Retrieval, gibt einen kurzen Überblick über bestehende Ansätze des Information Retrieval und erläutert die zunehmende Bedeutung benutzerverhaltensorientierter Relevanzbewertungs-Verfahren. Die geschätzte Größe des Internets wird dargelegt, um die zunehmende Bedeutung des Information Retrievals zu verdeutlichen, und bestehende Suchlösungen des Internets werden erläutert. Anschließend werden die Grundlagen für benutzerverhaltensorientierte Modelle näher erläutert.

2.1 Überblick über das Gebiet des Information Retrievals

Die Menge der verfügbaren Informationen im Internet steigt ständig an. Damit diese fast unüberschaubare Anzahl an Informationen in sinnvoller Weise zugänglich gemacht werden kann, bedarf es geeigneter Techniken, diese Informationen zu durchsuchen. Diese Techniken sollen es ermöglichen, für die Suchanfrage eines Nutzers relevante von nicht-relevanten Informationen unterscheiden zu können. [3,S.1]

Zur Entwicklung und Lösung von Suchtechniken gibt es verschiedene Ansätze im Feld des Information Retrievals *(IR)*. Obwohl es meist keine einheitliche Definition gibt, wird Information Retrieval in der Regel als das Gebiet definiert, welches die wissenschaftlichen Hintergründe und Forschungen für das Auffinden von Informationen in großen Datenbeständen liefert. „Wie der Begriff retrieval (deutsch Wiedergewinnung, Auffindung) sagt, sind Informationen in großen Datenbeständen zunächst verloren und müssen wiedergewonnen bzw. wiedergefunden werden.“ [7]

Dabei basieren die Prinzipien des IR meist auf den Forschungen in dem Bereich des Data Retrievals *(DR)* und überschneiden sich teilweise mit den dort existierenden Ergebnissen und Lösungen. Grundsätzlich können somit schon existierende Technologien aus dem Bereich des *DR* übernommen werden. Allerdings unterscheidet sich *IR* von *DR* in zwei wesentlichen Punkten:

1. Vagheit: “In data retrieval, we are normally looking for an exact match. In information retrieval this may sometimes be of interest but more generally we want to find those items which partially match the request and then select from those a few of the best matching ones.” [9] Es werden im *DR* exakte

Antworten auf eine Abfrage geliefert – Im *IR* werden nur verschiedene relevante Dokumente zu einer meist eher vagen Anfrage zurückgegeben. Dies wird durch eine grundlegende (und ältere) Definition des *IR* von Lancaster (1968) verdeutlicht: „An information retrieval system does not inform (i.e. change the knowledge of) the user on the subject of his inquiry. It merely informs on the existence (or non-existence) and whereabouts of documents relating to his request." [8]

2. Unsicherheit: Es fehlen dem System die notwendigen Kenntnisse über den Inhalt der Dokumente, da diese nicht nur als Text, sondern auch in Form von Bildern, Video etc. vorhanden sein können. Dies kann zu ungenauen Ergebnislisten der relevanten Dokumente führen - Probleme bei Texten bereiten z. B. Homonyme (Worte, die gleich geschrieben werden; z. B. Bank - Geldinstitut, Sitzgelegenheit) und Synonyme (Bank und Geldinstitut). [7]

Ein Überblick über den Bereich des *IR* gibt die folgende schematische Darstellung. (Abb. 1)

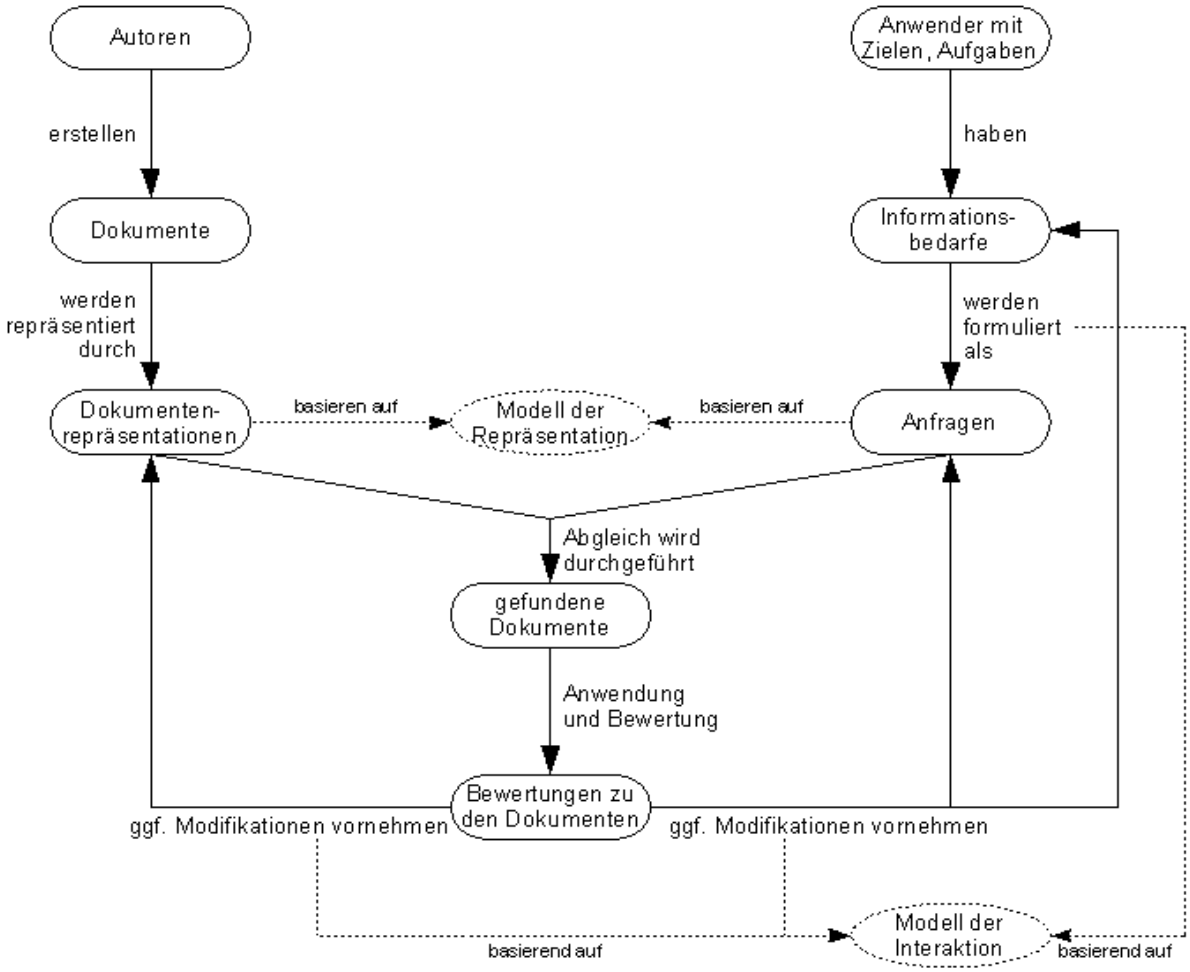


Abbildung 1: Schematisches Modell des Information-Retrieval [10]

Abbildung 1 verdeutlicht das Folgende: Es existieren zwei Personkreise; Autoren, welche Informationsobjekte in Form von Dokumenten und deren Repräsentationen erstellen und diese veröffentlichen, sowie Anwender (*Nutzer*), die Informationen in diesen Objekten suchen. Entsteht eine Suchanfrage eines Anwenders eines IR-Systems, so werden die bestehenden Dokumente mit der Abfrage abgeglichen und die gefundenen Dokumente bewertet und dem Anwender zurückgeliefert. Die Grundlage für das Abgleichen der Dokumente bildet ein Modell der Repräsentation von Dokumenten und Anfragen, über die beide vergleichbar gemacht werden.

Bedeutend für die weitergehende Betrachtung ist das so genannte IR-System, welches die Relevanzbewertung von den abgeglichenen Dokumenten anhand eines Gewichtungsalgorithmus vornimmt. „Zentrales Anliegen eines IR-Systems ist es, die Relevanz von Informationsobjekten (im Folgenden als Dokumente bezeichnet) in Hinblick auf eine Benutzeranfrage einzuschätzen. Eine solche Relevanz-Entscheidung wird von dem zugrunde liegenden Gewichtungsalgorithmus vorgenommen. Dabei etabliert der Gewichtungsalgorithmus eine Ordnung über den Dokumenten einer Kollektion, wobei gilt: Je höher ein Dokument gewichtet wird, desto höher ist dessen Relevanz für eine Benutzeranfrage einzustufen.“ [2,S.7]

Der Gewichtungsalgorithmus entscheidet darüber, welche Dokumente relevanter einzustufen sind als andere. Dies spielt bei der Übermittlung der Ergebnisse zum Anwender eine zentrale Rolle. Denn bei einer großen Menge an gefundenen relevanten Dokumenten vereinfacht die Sortierung der Dokumente nach Relevanz dem Anwender das Auffinden der gesuchten Informationen enorm.

2.2 *IR*-Modelle zur Relevanzbewertung

Existierende *IR*-Systeme unterscheiden sich in der Interpretation des Relevanzbegriffs. Dazu existieren unterschiedliche Modelle zur Einstufung der Relevanz und Gewichtung eines Objektes. So hängt die Relevanz zwischen Dokument und Anfrage zum einen von der jeweiligen Dokumentenrepräsentation und zum anderen von der verwendeten Strategie zur Abgleichung der Anfrage und der Dokumentenmenge ab.

Ein *IR*-Modell ist ein 4-Tupel $(D, Q, F, R(q,d))$, das wie folgt definiert ist: [2,7]

1. D ist eine Menge von Dokumenten einer Dokumentenmenge
2. Q ist eine Menge von Abfragen (Informationswünschen) eines Nutzers
3. F ist ein Rahmenwerk zur Modellierung von Beziehungen der Dokumente und Anfragen
4. $R(q,d)$ ist eine Retrievalfunktion, die zu einer gegebenen Anfrage $q \in Q$ eines Dokumentes $d \in D$ eine Zahl zuordnet. Diese als Gewichtung bezeichnete Zahl definiert eine Ordnung zwischen den Dokumenten in Bezug auf die Anfrage.

Es existieren verschiedene klassische Modelle zur Modellierung der Abgleichung eines Dokumentes (und seiner gegebenen Repräsentation) sowie der Anfragen. Wichtig sind unter anderem das Boolesche Modell und das Vektorraummodell. In beiden Modellen wird ein Dokument durch eine Anzahl von so genannten Schlüsselwörtern definiert, die als *Index-Terme* bezeichnet werden. Index-Terme sind Worte, die ein Dokument beschreiben und dessen Inhalt möglichst treffend charakterisieren. Der Vorgang des Zuordnens von Index-Termen zu Dokumenten wird als Indexierung bezeichnet. Die Zuordnung erfolgt in der Regel automatisiert, kann aber auch manuell erfolgen.

Das Boolesche Modell zeichnet sich durch seine Einfachheit aus: Dokumente werden als relevant eingestuft, wenn der Suchbegriff in dem Dokument vorkommt. Dabei kann die Häufigkeit der Vorkommnisse, die relative Position der Vorkommen untereinander, die Struktur des Dokumentes sowie eine benutzerdefinierte Gewichtung des Suchbegriffes mit berücksichtigt werden.

Gleiche Index-Terme können dabei in unterschiedlichen Dokumenten unterschiedlich bedeutsam sein. So kann z.B. der Index-Term „Flugzeug“ für Dokumente im Bereich Flugzeugkonstruktion bedeutender sein als für Dokumente, die sich allgemein mit Transportmitteln beschäftigen. Deswegen kann jedem Index-Term eines Dokumentes zusätzlich ein numerisches Gewicht zugeordnet werden. Das Vektorraummodell arbeitet mit Gewichtungen der Index-Terme des Dokumentes. Die Terme eines jedes Dokumentes werden nach einer bestimmten Gewichtungs-Methode gewichtet. Ebenfalls werden die Terme der Anfrage gewichtet und diese

dann mit den jeweils gewichteten Termen des Dokumentes verglichen. Das Vektorraummodell kommt dabei ganz ohne Boolesche Operatoren aus und liefert so eine tolerantere Interpretation des Relevanzbegriffes als das Boolesche Modell. [2]

Weiterhin gibt es noch Ansätze, die im Bereich des sog. „Web-Mining“ liegen. Diese Ansätze betrachten die Verlinkungsstruktur der Dokumente untereinander und folgern aus dieser Struktur, zu welchen Anfragen die Seite sich als relevant erweisen kann. Bekanntes Beispiel hierzu ist das so genannte „PageRank“-Verfahren. Des Weiteren gibt es noch verschiedene Ansätze, die Web-Seite in verschiedene Klassen ihrer Größe zu unterteilen. So werden Dokumente anhand ihrer Verlinkung in sog. „Authorities“ und „Hubs“ unterteilt, wobei „Authorities“ z.B. Portalseiten und „Hubs“ kleinere Seiten darstellen. [2]

Im Folgenden werden das Boolesche Modell und das Vektorraummodell genauer betrachtet.

2.3 Boolesche Retrieval - Modell

Das Boolesche Retrieval basiert auf der Mengentheorie und den recht bekannten booleschen Operationen „und“, „oder“ und „nicht“ $(\wedge, \vee, \neg)$. Benutzeranfragen bestehen aus Termen, die mittels dieser Operationen nach den folgenden Regeln miteinander verknüpft werden können (Mit Q als Menge der Abfragen und T als Menge der Index-Terme eines Dokumentes):

1. $t_i \in T \rightarrow t_i \in Q$ bedeutet, dass ein Index-Term eines Dokumentes auch als Abfrage formuliert werden kann.
2. $q_1, q_2 \in Q \rightarrow q_1 \wedge q_2 \in Q$ bedeutet, dass ein Index-Term einer Abfrage mit der Operation $\wedge$ („und“) verknüpft werden kann.
3. $q_1, q_2 \in Q \rightarrow q_1 \vee q_2 \in Q$ bedeutet, dass ein Index-Term einer Abfrage mit der Operation $\vee$ („oder“) verknüpft werden kann.
4. $q \in Q \rightarrow \neg q \in Q$ bedeutet, dass ein Index-Term einer Abfrage mit der Operation $\neg$ („nicht“) negiert werden kann.

Es wird davon ausgegangen, dass ein Index-Term in einem Dokument entweder vorhanden oder nicht vorhanden ist. Somit liefert die zugehörige Retrievalfunktion

$R(q_i, d_i)$ nur den Wert 1 (Dokument ist relevant) und den Wert 0 (Dokument ist nicht relevant) zurück.

Die Verwendung der Negation wird dabei in der Regel eingeschränkt, so dass sie nur in Kombination mit Konjunktionen verwendet werden darf. Somit sind Anfragen wie $q = t_1 \vee \neg t_2$ oder $\neg t_1$ ungültig, da sie für den Anwender meist nicht nachvollziehbare Anfragen formulieren.

Der Nachteil beim Booleschen Retrieval ist, dass die Größe der Antwortmenge stark variieren kann. Wird eine zu restriktive Anfrage gestellt, so werden nur wenige Dokumente gefunden. Ist die Anfrage zu generell, kann die Antwortmenge auf ein unübersichtliches Maß anwachsen. Ein weiteres Problem besteht darin, dass beim Booleschen Retrieval keine Ordnung auf den Dokumenten der Antwortmenge definiert ist, um eine Sortierung vorzunehmen. Deswegen werden meist nachträglich zusätzliche Kriterien verwendet, um eine Sortierung nach Relevanz der Dokumente herzustellen:

- Je häufiger der Index-Term in dem Dokument vorkommt, umso relevanter ist das Dokument
- Der Nutzer kann bei der Suchanfrage eine Gewichtung der Index-Terme vorgeben, die bei der Sortierung berücksichtigt werden soll
- Die relative Position der Index-Terme innerhalb des Dokuments wird berücksichtigt – je näher, umso relevanter.
- Die Semantik bzw. die Dokumentstruktur wird berücksichtigt – Dokumente, deren Überschrift den Index-Term enthalten, können relevanter sein als Dokumente, in denen der Index-Term nur im Textkörper vorkommt

2.4 Vektorraummodell

Die im Booleschen Modell durchgeführte Trennung in eindeutig relevante und nicht-relevante Dokumente kann als zu restriktiv angesehen werden. Bei einer Nutzeranfrage (z.B.: $q = t_1 \wedge t_2 \wedge t_3$) werden neben Dokumenten, die keinen der Index-Terme enthalten, auch Dokumente abgewiesen, die einen oder zwei der Index-Terme enthalten. Auch das Problem der Formulierung von booleschen Anfragen bereitet der Masse von Nutzern Probleme. Deswegen wurde im Rahmen des experimentellen IR-Systems SMART [18] das Vektorraummodell entwickelt.

Im Vektorraummodell wird der Begriff der Relevanz toleranter formuliert, da es ohne boolesche Operatoren auskommt. Es wird ein nicht-binärer Wert für die Gewichtung der Index-Terme der Anfrage und der Dokumente verwendet. Dazu ist jedem Paar eines Index-Terms und eines Dokumentes (t_i, d_j) ein positives, nicht binäres Gewicht $w_{i,j}$ zugeordnet. Weiterhin ist jeder Index-Term t_i einer Suchanfrage q mit dem Gewicht $w_{i,q}$ gewichtet. Die komplette Anfrage kann durch einen Vektor dargestellt werden: $\vec{q} = (w_{1,q}, w_{2,q}, ..., w_{n,q})$ mit n als Anzahl der Gesamt-Index-Terme der Suchanfrage. Der Vektor der Gewichtungen der Index-Terme eines Dokumentes d_j ist definiert durch $\vec{d}_j = (w_{1,j}, w_{2,j}, ..., w_{n,j})$ Mit einem gebräuchlichen Ähnlichkeitsmaß für Vektoren, wie z.B. dem Kosinus-Maß, kann der Grad der Korrelation zwischen einer Anfrage und einem Dokument bestimmt werden. Anhand der Ähnlichkeit kann somit eine Ordnung der Dokumente nach der Relevanz erstellt werden.

2.4.1 Gewichtungsmethoden im Vektorraummodell

Das Hauptproblem im Vektorraummodel ist die automatisierte Gewichtung der einzelnen Terme in den Anfragen sowie in den Dokumenten. Dabei wird zwischen lokalen und globalen Gewichten unterschieden. Lokale Gewichtungen beachten dabei z.B. die Häufung von Index-Termen innerhalb eines Dokumentes, globale Gewichtungen die Häufung z.B. innerhalb der Sprache.

2.4.2 Globale Gewichtung im Vektorraummodell

Die Verteilung der Wörter in einer Sprache kann grob durch das Zipf'sche Gesetz beschrieben werden. [19] Aus diesem Gesetz kann abgeleitet werden, dass eine kleine Anzahl von häufigen Wörtern einen großen Anteil an Texten abdeckt, und eine große Anzahl an seltenen Wörtern nur einen kleinen Teil eines Textes ausmachen. Häufige Terme sind somit keine geeigneten Index-Terme, da mit Ihnen der Text nicht spezifisch genug beschrieben werden kann. Sie sollten eher in Ihrer Bedeutung durch eine niedrigere Gewichtung abgeschwächt werden oder als sog. Stoppwort deklariert werden. (Stoppworte sind Worte, die sehr häufig in der Sprache vorkommen und deswegen aus Dokumenten und Anfragen herausgefiltert werden, z.B. „der“ oder „und“).

In der Regel wird die „inverse Dokumenthäufigkeit“ (*inverse document frequency, IDF*) für einen Term verwendet, um einen häufigen Term nicht so stark zu gewichten wie einen seltenen. Die *IDF* wird meist berechnet mit

$$IDF = \log\ (\ N / n\ )$$

mit *N* = Anzahl an Dokumenten im System und *n* = Anzahl an Dokumenten, die den Term beinhalten. [6, S.198]

2.4.3 Lokale Gewichtung

Bei der lokalen Gewichtung wird die Häufigkeit von Termen (*term frequency, tf*) innerhalb der einzelnen Dokumente zur Bestimmung der Gewichtung betrachtet. Terme, die häufiger in einem Dokument auftauchen, sind bessere Deskriptoren für ein Dokument als selten vorkommende.

Um die unterschiedliche Länge von Dokumenten auszugleichen, wird meist die Häufigkeit eines Terms t_i in einem Dokument d_j zum Maximum aller Termhäufigkeiten in diesem Dokument in Bezug gesetzt. Es gilt somit für die normalisierte Termhäufigkeit $ntf_{i,j}$:

$$ntf_{i,j} = \frac{tf_{i,j}}{\max_{l \in \{1..n\}} tf_{l,j}}$$

Dabei ist $tf_{i,j}$ die absolute Häufigkeit eines Terms t_i in einem Dokument d_j. [2, S.9f]

2.4.4 *tf-idf* Gewichtung

Durch die Kombination von lokaler und globaler Gewichtung erhält man die sog. *tf-idf*-Gewichtung. Die Gewichtung $w_{i,j}$ der einzelnen Terme von Dokumente sowie einer Anfrage können durch die Multiplikation der normalisierten Termhäufigkeit mit der inversen Dokumentenhäufigkeit berechnet werden:

$$w_{i,j} = ntf_{i,j} \cdot idf_i$$

2.5 Metriken zur Bestimmung der Retrievalqualität

Im Bereich des *IR* gibt es viele Ansätze, um konkrete Aussagen über die Qualität bzw. Quantität eines *IR*-Systems machen zu können. Allerdings machen solche Metriken nicht immer sinnvolle Aussagen über die Retrievalqualität, da sie meist

bestimmte Faktoren „außen vor“ lassen: „Much effort and research has gone into solving the problem of evaluation of information retrieval systems. However, it is probably fair to say that most people active in the field of information storage and retrieval still feel that the problem is far from solved.” [9,S.144] Einer dieser Faktoren ist z.B. die Subjektivität des Relevanzbegriffes – die Relevanz eines Dokumentes kann je nach Nutzer und dessen momentanen Infomationswunsch variieren. [9,S.147]

Es werden nun die drei in der Literatur bekanntesten Metriken im *IR* vorgestellt: Die Abdeckung (*Recall*), die Präzision (*Precision*) und der Ausfall (*Fallout*), da mit Ihnen meist sinnvolle Angaben über ein Retrievalsystem gemacht werden kann.

Zur Erklärung dieser drei Metriken ist die folgende Tabelle der Abhängigkeiten von gefundenen und relevanten Dokumenten hilfreich. (Tab. 1) *A* bezeichnet dabei relevante Dokumente, *B* bezeichnet gefundene Dokumente. Entsprechend stehen $\overline{A}$ für nicht-relevante und $\overline{B}$ für nicht gefundene Dokumente. [9,S.148]

	Relevant	Nicht-relevant	
Gefunden	$A \cap B$	$\overline{A} \cap B$	B
Nicht gefunden	$A \cap \overline{B}$	$\overline{A} \cap \overline{B}$	$\overline{B}$
	A	$\overline{A}$	N

N ist die Menge aller Dokumente im System.

Tabelle 1: Abhängigkeiten von gefundenen und relevanten Dokumenten

Die Abdeckung (*Recall*) beschreibt die Vollständigkeit eines Suchergebnisses. Sie ist definiert als der Anteil bei einer Suche gefundenen relevanten Dokumenten aus den relevanten Dokumenten der Grundgesamtheit. [7] Mit ihm kann ausgesagt werden, wie viele relevante Dokumente auch gefunden werden.

$$\mathrm{Re}call = \frac{|A \cap B|}{|B|}$$

Die Präzision (*Precision*) beschreibt die Genauigkeit eines Suchergebnisses. Sie ist definiert als der Anteil relevanter Dokumente von allen bei einer Suche gefundenen Dokumenten. [7] Mit ihr wird ausgesagt, wie viele der gefundenen Dokumente relevant sind.

$$\Pr ecision = \frac{|A \cap B|}{|A|}$$

Der Ausfall (*Fallout*) beschreibt, wie viele der nicht-relevanten Dokumente gefunden wurden. [9,S.148f.]

$$Fallout = \frac{|\overline{A} \cap B|}{|\overline{A}|}$$

Die Maße der Abdeckung und der Präzision sind einander gegenläufig. Dies lässt sich anschaulich durch zwei Extremfälle verdeutlichen:

1. Wenn alle Dokumente auf eine Anfrage hin zurückgeliefert werden, ist die Abdeckung gleich 1, da sich natürlich auch alle relevanten Dokumente unter der Antwortmenge befinden. Da die Antwortmenge damit aber fast nur aus irrelevanten Dokumenten besteht, wird die Präzision dementsprechend niedrig ausfallen.
2. Wird umgekehrt nur ein einziges relevantes Dokument gefunden, so ist die Präzision gleich 1, aber die Abdeckung wird dementsprechend niedrig ausfallen, da die meisten relevanten Dokumente nicht gefunden wurden.

Ein mathematischer Beweis dieser Gegenläufigkeit ist möglich, würde aber im Rahmen dieser Diplomarbeit zu weit gehen. Er findet sich in [9,S.149].

2.6 Größe des Internets

Insbesondere das Internet ist als ein sehr großer Datenbestand anzusehen. Letzte verlässliche Zahlen und Studien geben an, dass Anfang 2001 über 2 Milliarden Webseiten existieren. [13] Im Dezember 2002 waren es schon über 3 Milliarden. [14] Diese Zahlen basieren auf der Größe des Index von bekannten Suchmaschinen. Hierbei nicht eingeschlossen ist das so genannte „Deep Web“, welches von den Suchmaschinen nicht indexierbare Seiten beinhaltet (aufgrund fehlender Verlinkung bzw. dynamischen Inhalts dieser Seiten) und somit schwer mit einbezogen werden kann. Das „Deep Web“ soll nach Schätzungen noch bis zu 550-mal größer sein als der den Suchmaschinen zugängliche Teil des Webs. [15]

Aktuelle Zahlen (Anfang 2005) geben an, dass es schätzungsweise 12,5 Milliarden Webseiten gesamt geben soll. Dabei hat Google ca. 8 Milliarden und Yahoo ca. 4 Milliarden Seiten im Index. [16] Google ist dabei „eines der größten Rechenprojekte der Welt, das wohl mehr Rechner einsetzt als jedes andere voll verwaltete Einzelsystem (...)“ [4], was zeigt, dass für das Indexieren und Durchsuchen dieser großen Mengen viel Rechenleistung notwendig ist.

Interessant sind diese Zahlen neben ihrer unvorstellbar großen Dimension auch für *IR*-Systeme: „Sind Größe und Wachstum bekannt, ermöglicht dies z. B. vergleichende Aussagen über den Grad der Abdeckung des WWW durch die verschiedenen generellen Web-Suchmaschinen wie Google (..)“ oder Yahoo. Des Weiteren wird durch diese Zahlen die zunehmende Bedeutung von IR-Systemen für das Auffinden von Informationen deutlich. Im Folgenden wird ein Überblick über bestehende *IR*-Systeme im Internet gegeben.

2.7 Bestehende *IR*-Systeme im Internet

Als *IR*-Systeme sind Suchmaschinen und Verzeichnisse im Internet anzusehen. Bekannte Beispiele hierfür sind u.a. Google[2] als automatisierte Suchmaschine und das Open Directory[3] als reiner Verzeichnisdienst sowie Yahoo[4] als Mischung eines Verzeichnisdienst und eines automatisierten Suchdienst.

Es lassen sich zur Zeit drei Klassen von typischen Suchlösungen im Internet spezifizieren: Kataloge, Suchmaschinen und Meta-Suchmaschinen. Weiterhin werden alternative Suchansätze betrachtet, insbesondere Suchdienste, die ein sog. „benutzerverhaltenorientiertes (*personalisiertes*) Profil“ des Anwenders berücksichtigen.

2.7.1 Kataloge

Kataloge im Internet basieren auf einer Hierarchie von vordefinierten Schlagworten, den so genannten Kategorien. Zum Hinzufügen von neuen Webseiten werden diese redaktionell bearbeitet und von menschlichen Indexiern den einzelnen Kategorien zugeordnet. Innerhalb dieser ausgewählten Kategorien werden weiterhin die

[2] http://www.google.de
[3] http://www.dmoz.org
[4] http://www.yahoo.de

entsprechenden Links zu dieser Web-Seite sowie eine kurze Inhaltsbeschreibung bereitgestellt. Die neu zu erfassenden Web-Seiten werden den Betreibern der Katalogdienste meist von den Autoren selbst mitgeteilt. Es werden von den Katalogdiensten aber auch eigenständige Crawler-Programme zum Aufspüren neuer, populärer Web-Seiten betrieben.

Katalogdienste führen durch Ihre einfache Navigation und der hohen Qualität der thematischen Relevanz der Dokumente meist schnell zu den nötigen Informationen einer Anfrage. Allerdings stellt die Kategorisierung der Web-Seiten durch menschliche Fachkräfte einen arbeitsintensiven Prozess dar und kann deshalb nur schwer mit dem Wachstum des Internets Schritt halten. Was bedeutet, dass die Abdeckung der Katalogdienste recht gering ist.

Zudem bringt die Auswahl der Kategorien Probleme mit sich: „Die Konstruktion der Hierarchie erscheint als einigermaßen hybrides Projekt, zielt es doch darauf ab, Millionen völlig heterogener Netzbeiträge aus nahezu allen Bereichen der menschlichen Wissensbestände auf ein einheitliches Kategoriensystem zu bringen, ungeachtet ihrer Perspektive, ihrer Widersprüche und Konkurrenzen.“ [11] Es können bestimmte Themengebiete nur bedingt einzelnen Kategorien zugeordnet werden – beispielsweise kann die Kategorie Umweltverschmutzung als Unterkategorie von Gesellschaft sowie als Unterkategorie von Natur eingeordnet werden. Weiterhin können Dokumente zum Thema „Medieninformatik“ mehr als nur einer Kategorie angehören. Hinzu kommt, dass durch unterschiedliche subjektive Auffassungen der Nutzer über den erwarteten Inhalt ein intuitives Zurechtfinden in diesen Kategorien erschwert wird.

Beispiele für einen Katalogdienst sind Yahoo und das Open Directory. Als einer der ersten Suchdienste im Internet verlieren Kataloge nach und nach aufgrund der stark wachsenden Anzahl an Webseiten und der damit verbundenen schwierigen Pflege ihre Bedeutung. "We analyzed what people were using, and that [the Open Directory] had become less popular over time. As the web grows, directory structures get harder to use." [20]

2.7.2 Suchmaschinen

Die zentralen Teile einer Suchmaschine sind zum einem der so genannte *Crawler* (auch *Robot* oder *Spider* genannt) sowie der Gewichtungsalgorithmus zur Treffersortierung.

Der Crawler durchsucht das Internet (durch systematische Verfolgung von Links oder durch das direkte Besuchen einer angemeldeten Webseite) nach neu entdeckten bzw. aktualisierten Seiten. Der Inhalt dieser Seite wird an den Server übermittelt, welcher die entsprechenden Index-Terme ermittelt und diese dem Index der Suchmaschine hinzufügt.

Eine Anfrage des Nutzers erfolgt in der Regel über eine webbasierte Schnittstelle in Form von Suchtermen. Dabei können diese Suchterme durch logische bzw. selbst definierte Operationen verknüpft sein (meist durch eine logische UND-Verknüpfung). Anhand dieser Anfrage wird eine Ergebnisliste zurückgeliefert und zur Anzeige gebracht. Dabei werden die Dokumente zuerst angezeigt, deren Relevanz nach dem Gewichtungsalgorithmus der Suchmaschine am höchsten bewertet wurde.

Dokumente werden durch den Link auf das Originaldokument sowie einem kurzen beschreibenden Text – einem *Snippet* – repräsentiert. Dieser Snippet ist in der Regel automatisch generiert und besteht aus dem Titel des Dokumentes, eventuellen vorhandenen Metainformationen sowie Auszügen aus dem Text des Dokumentes.

Das Geheimnis einer guten Suchmaschine liegt in dem verwendeten Algorithmus zur Relevanzbewertung. Dabei verwenden Suchmaschinen meist Varianten des Booleschen Modells und des Vektorraummodells zum Auffinden der relevanten Dokumente. Im Laufe der Zeit sind diese Verfahren aber auch an ihre Grenzen angelangt, sodass sich Verfahren wie z.B. das „PageRank“ mit dem bekannten Beispiel Google durchgesetzt haben. Das PageRank-Verfahren berücksichtigt dabei die Verlinkungsstruktur der Dokumente untereinander.

2.7.3 Alternative Suchansätze

Die zwei oben genannten Suchlösungen, Katalogdienste und Suchmaschinen, verhalten sich gegensätzlich gegenüber ihrer Abdeckung und Präzision. Die Anzahl

an indizierten Webseiten von Suchmaschinen umfasst einen großen Teil der existierenden Dokumente – Suchmaschinen gewährleisten also einen hohen Grad der Abdeckung. Allerdings besitzen sie aufgrund der oft sehr großen Ergebnislisten eine geringe Präzision, da die relevanten gefundenen Seiten in der Masse der irrelevanten gefundenen Seiten nur schwer zu finden sind. Im Gegensatz dazu bieten Kataloge Links zu Webseiten von hoher Qualität und großem Themenbezug, haben also einen hohen Präzisionswert. Kataloge decken aber nur einen kleineren Teil der tatsächlich im WWW vorhandenen Seiten zu einem bestimmten Thema ab.

Aufgrund dieses Unterschieds sind Suchlösungen entstanden, die diesem Problem Rechnung tragen. Ziel dieser Lösungen ist, einen hohen Wert der Präzision und gleichzeitig einen höheren Wert der Abdeckung zu erreichen. Inzwischen bieten viele Suchdienste im Netz deshalb Lösungen an, die parallel sowohl einen Katalog als auch eine Suchmaschine bereitstellen.

Allerdings existiert bei allen vorhandenen Suchmaschinen ein Problem bei der Sortierung der Ergebnislisten. Das Problem besteht darin, die relevanteren Treffer in der Ergebnisliste von den wahrscheinlich irrelevanteren Treffern abzuheben. Dies wird von vielen Suchmaschinen unterschiedlich gelöst, meist aber nicht zur vollständigen Zufriedenheit des Nutzers. Deswegen gibt es neben den Katalogen und den Suchmaschinen noch alternative Suchdienste:

- *Natürlichsprachige Suchdienste*: Bei dieser Art von Suchdienst werden Ansprachen natürlichsprachig formuliert. Der Suchdienst liefert daraufhin Seiten zurück, die die Frage exakt beantworten. Fragen und Antwortseiten werden durch redaktionelle Bearbeitung einander zugeordnet. Somit stellen natürlichsprachige Suchdienste nur eine Sonderform der Web-Kataloge dar. Bekanntes Beispiel hierfür ist der Suchdienst AskJeeves[5].

- *Metasuchdienste:* Diese Suchdienste setzen auf anderen Suchmaschinen auf und fragen diese ab. Erreicht eine Anfrage eine Metasuchmaschine, so fragt diese die anderen Suchmaschinen ab, wertet dessen Antwortmenge aus

[5] http://www.aksjeeves.com

(durch Zusammenfassen und evtl. neu sortieren) und liefert die Ergebnisse zurück.

- *Benutzerverhaltensorientierte Suchdienste*: Diese Suchdienste betrachten die persönlichen Vorlieben eines Nutzers und berücksichtigen diese bei der Sortierung der Antwortmenge.

Im Folgenden wird auf die Metasuchmaschinen sowie auf die benutzerverhaltensorientierten Suchdienste detaillierter eingegangen. Die natürlichsprachigen Suchdienste als Sonderform der Web-Kataloge wurden nur der Vollständigkeit halber erwähnt und werden nicht weiter erläutert.

2.7.4 Metasuchmaschinen

Studien zeigen, dass der Suchraum durch das Kombinieren von Suchmaschinen erweitert werden kann. Die Suchlösung wird durch so genannte Meta-Suchmaschinen umgesetzt. Über eine Suchschnittstelle werden die Ergebnisse verschiedener Suchmaschinen zusammengefasst, verglichen und die Treffer erneut sortiert. Allerdings tritt bei diesen Suchmaschinen meist das Problem einer längeren Antwortzeit (meist 5 bis 10 Sekunden) auf, da das Abfragen und Vergleichen mehrerer Suchmaschinen aus verschiedenen Gründen ein zeitintensiver Prozess ist.

Bei Metasuchdiensten existieren verschiedene Abfragetechniken der Suchergebnisse. Typische Metasuchdienste stellen die vereinten Suchergebnisse aller abgefragten Suchdienste dar, ohne dabei die Quelle der Information mit einzubeziehen. Die Relevanz der einzelnen Dokumente wird durch die Rangfolge (*Ranking*) in den jeweils abgefragten Suchdiensten ermittelt. In fortgeschritteneren Verfahren werden nach Abfrage der Ergebnisliste von den Suchmaschinen zudem noch die Dokumente selber betrachtet und so eine erweiterte Relevanzbewertung unter Berücksichtigung der jeweiligen Quelle vorgenommen. [12]

2.7.5 Benutzerverhaltensorientierte Suchdienste

Es gibt Hinweise darauf, dass existierende Suchmaschinen nicht den Anforderungen ihrer Nutzer gewachsen sind. [5] „Every day millions of people trawl the Internet for information using any one of a dozen or more different search tools – but whether

they find what they are looking for sometimes depends not only on their skills, but also on their luck." [6,S.194]

Dieses Problem resultiert vor allem aus dem Umstand, dass der Nutzer die Anfrage meist nicht korrekt formulieren kann und die Antwort auf eine Suchabfrage somit nicht die gewünschten Ergebnisse liefert. „Most problems centre around difficulties in finding relevant information. Simple keyword queries often yield too many results of variable quality. It is usual to receive several hundred hits (documents matching the user's query) even on relatively narrow queries and to receive thousands of hits for a query is not uncommon. Research also shows that users do not find it easy to frame the query needed to return the information they require." [6,S.194]

Weiterhin wird beschrieben, dass viele Nutzer zwar mit dem bestehenden System des Suchens zufrieden sind, nämlich dem Beschreiben ihres Infomationswunsches durch ein oder mehrere Schlüsselwörter. Anderseits ist dem Nutzer nicht bewusst, dass die Angabe eines kompletten Interessensgebiets bei der Suchanfrage ggf. bessere Ergebnisse liefern würde. Obwohl der Nutzer es z.B. gewohnt ist, bei einer Recherche in einer Bibliothek nach einem bestimmten Buch dem Bibliothekar ganze Themenfelder zu beschreiben, wird dieses Verhalten bei der Suche im Internet nicht so umgesetzt. Dort werden nach Angabe von wenigen Schlüsselwörtern meist exakte Ergebnisse erwartet Und selbst wenn komplexere Anfragen vom Nutzer formuliert werden möchten, bieten die gängigen Suchmaschinen nicht die entsprechenden Eingabemöglichkeiten.

Das Hauptproblem liegt somit darin, dass der Anwender seine Anfrage durch kurze Stichwörter charakterisieren muss. „Eine mechanische Stichwortsuche setzt voraus, dass nur solche Fragen gestellt werden, die in Stichworten klar formulierbar sind und durch weitere Stichworte differenziert und konkretisiert werden können. Ebenso wird niemand erwarten, dass das System neben dem gefragten auch bedeutungsähnliche Begriffe einbeziehen oder Homonyme ausschließen kann. Alle Fragen, die auf Stichworte nicht zu reduzieren sind, fallen aus dem Raster des Möglichen heraus; technische und naturwissenschaftliche Termini werden sich relativ gut für die Suche eignen, geisteswissenschaftliche Themen weit weniger gut." [11]

Der Nachteil, die Anfrage des Anwenders ausschließlich durch kurze und prägnante Terme zu formulieren, kann durch ein schon eingangs erwähntes Beispiel verdeutlicht werden. Die Abfrage mit dem Suchbegriff “Apple” eines englischsprachigen Nutzers kann sich auf Dokumente beziehen, die sich mit der Frucht befassen, während ein anderer Nutzer Dokumente sucht, die sich mit dem gleichnamigen Computer-Hersteller thematisch auseinandersetzen. [1,S.2]

Der Ansatz in benutzerverhaltensorientierten (sog. *personalisierten*) Suchdiensten besteht darin, die bisherigen Anfragen an Suchdienste um Informationen über den Nutzer zu erweitern. Stellt der Nutzer eine Suchabfrage an einen Suchdienst, so werden die Vorlieben und Interessen des Nutzers (das sog. *Profil des Nutzers*) ebenfalls bei der Ermittlung der Antwortmenge berücksichtigt. Somit können bei einer Suchabfrage, die wie oben beschrieben meist aus wenigen Stichwörtern besteht, weitere Informationen verarbeitet und die ermittelten Ergebnisse an die persönlichen Vorlieben und Interessen des Nutzers angepasst werden.

Das Gebiet der personalisierten Suchdienste ist noch in Ihren Anfängen. Es gibt zurzeit zwei verschiedene Ansätze, die Lösungen in diesem Bereich beschreiben [6]:

- inhaltsbasierende Personalisierungsansätze, die ein erstelltes Profil mit den Inhalt des Dokumentes abgleichen
- soziale Filterung („*social filtering*“), die eine Beziehung zwischen den Betrachtern eines Dokumentes herstellt

Beiden Ansätzen ist gemein, dass ein Profil erstellt werden muss, welches die Vorlieben und Interessen des Nutzers charakterisiert. Die Erstellung sowie der Abgleich des Profils mit den jeweiligen Dokumenten stellt eines der Hauptprobleme aller personalisierten Suchtechniken dar.

Zudem unterstützen und erweitern die personalisierten Suchlösungen die bisher klassischen Suchdienste – es ist immer eine formulierte Anfrage von Schlüsselwörtern des Nutzers nötig. Lediglich werden bei der Anfrage noch weitere Informationen über den Nutzer hinzugezogen, um die Antwortmenge besser auf die Vorlieben des Nutzers anpassen zu können

2.7.5.1 Inhaltsbasierende Personalisierungs-Ansätze

Diese Ansätze basieren auf den in 2.3 und 2.4 beschriebenen Booleschen Modell bzw. Vektorraummodell. Dazu wird vereinfacht das Boolesche Modell als Vektorraummodel mit den jeweils zulässigen Gewichten 1 oder 0 angenommen.

Über ein automatisches oder manuelles Verfahren wird ein Profil des Nutzers erstellt. Diese Profilerstellung liefert einen Vektor, in dem die Vorlieben eines Nutzers gespeichert werden. Erreicht nun eine Suchanfrage eines Nutzers den Suchdienst, so wird ebenfalls der Vektor des Profils des Anwenders mit übermittelt und bei der Gewichtung der Dokumente nach den im Vektorraummodell beschriebenen Abgleichverfahren mit berücksichtigt.

Das Hauptproblem in dieser Lösung besteht in der Erstellung des persönlichen Profils. Dieses Profil kann manuell erstellt werden, welches aber aus vernünftigen Annahmen heraus schwer umzusetzen ist. Das automatisierte Erstellen ist somit erforderlich, bringt aber einige Probleme mit sich. Für die automatisierte Erstellung solch eines Profils gibt es verschiedene Ansätze, von denen drei beispielhaft vorgestellt werden.

Ein Ansatz ist die Erstellung eines Agenten, welcher die betrachteten Dokumente eines Nutzers auswertet, den Nutzer ggf. um ein Feedback zu diesem Dokument bittet und einen Vektor mit den Index-Termen und dem aus dem Feedback ermittelten Gewicht jedes betrachteten Dokumentes an den Suchdienst übermittelt. Aus den übermittelten Daten ergibt sich eine persönliche Gewichtung der Suchterme des Nutzers, die auf einer Auswertung der betrachteten Dokumente basieren. [6,S.199]

Ein weiterer Ansatz für die automatisierte Erstellung eines persönlichen Profils wird in [17] beschrieben. Dort wird das Problem mit Hilfe von Linguistischen Mustern gelöst. Es werden alle Dokumente des Nutzers betrachtet, die der Nutzer als „von ihm selbst verfasst“ oder „gerne gelesen“ einstuft. Diese Dokumente werden dann mit Hilfe von linguistischen Mustern analysiert und aus diesen das persönliche Profil des Nutzers erstellt. Dokumente, die ähnliche linguistische Muster enthalten wie die im Profil des Nutzers angegebenen, werden als für den Nutzer interessant eingestuft.

Ein Beispiel für eine Profilerstellung mit Hilfe eines Neuronalen Netzwerkes wird in [6,S.200] gegeben. Dort wird ein Netzwerk beschrieben, welches das Profil anhand der aufgerufenen Dokumente des Nutzers beobachtet und daraus selbstständig Schlussfolgerungen zieht. Betrachtet der Nutzer z.B. viele Dokumente, welche mit dem Thema „Telekommunikation“ verwandt sind, und ein unbekanntes Dokument beschäftigt sich mit „Telefon Kommunikation“, so kann das Neuronale Netz eine Relevanz dieses Dokumentes einschätzen, da die Terme „Telekommunikation, „Telefon“ und „Kommunikation“ im Netz untereinander Verbindungen und entsprechende Gewichtungen besitzen.

Der Vorteil der inhaltsbasierten Personalisierungs-Ansätze ist ihre Adaption von einfachen und effizienten Techniken aus dem Bereich des *IR*, wie z.B. dem Vektorraummodell. Der Nachteil besteht darin, dass sie ausschließlich für Textdokumente angewendet werden können. Zudem gibt es Probleme mit Homonymen und Synonymen, die im Profil vorkommen können. Dies kann durch eine Verbesserung der Profilerstellung vermieden werden, indem bestimmte Index-Terme des Profils erweitert werden. So werden den Index-Termen des Profils die Synonyme der Terme hinzugefügt, sprachliche Hierarchien und Abhängigkeiten dieser Terme berücksichtigt und Bezüge zu der Abfrage hergestellt.

Ein Beispiel für die inhaltsbasierte Personalisierung von Suchergebnissen findet sich in den Google Labs[6], in denen Beta-Produkte von Google vorgestellt werden.

2.7.5.2 Soziale Filterung

Die oben beschriebenen Ansätze basieren ausschließlich auf der Analyse von Schlüsselwörtern in Dokumenten und Profilen. Eine weitere Vorgehensweise ist das als soziale Filterung bekannte Verfahren, welches die Korrelation zwischen den verschiedenen Nutzern von Dokumenten beobachtet und bewertet.

Das Prinzip der sozialen Filterung geht von der Annahme aus, dass Dokumente und Informationen, die für eine Person in einer Gemeinschaft von Interesse sind, wahrscheinlich auch für andere Personen dieser Gemeinschaft von Interesse sind.

[6] http://labs.google.com/personalized

Die soziale Filterung begründet sich nicht auf die Beziehung von Informationen, sondern auf die Beziehung von Nutzern dieser Informationen. [21]

Als Beispiel für ein Konzept der sozialen Filterung ist [22]. Es werden Nutzer beim Bearbeiten Ihrer e-Mails zunächst beobachtet und deren Aktivitäten protokolliert. Nach einer gewissen Zeit kann das System dann den Nutzern sinnvolle Vorschläge zur Bearbeitung neuer, unbekannter e-Mail-Eingänge machen. Grundlage dieser Vorschläge ist das Verhalten der Nutzer bei bestimmten Inhalten der e-Mails. Erhält ein anderer Nutzer eine e-Mail mit ähnlichen Inhalt, so wird ihm ein Verhalten vorgeschlagen, dass bereits ein oder mehrere ihm ähnliche Nutzer durchgeführt haben.

Soziale Filterung berücksichtigt somit nicht die Inhalte von Dokumenten, sondern die Nutzer dieser Dokumente und deren Beziehungen untereinander. Der nun folgende Teil dieser Arbeit beschäftigt sich mit der Entwicklung eines Algorithmus, welcher auf den Grundlagen des Prozesses der sozialen Filterung basiert und eine Bewertung einer Dokumentenmenge auf Basis einer Beziehung zwischen den Nutzern von Dokumenten vornimmt.

Im Weiteren wird die Suchmaschine „Fooxx“ vorgestellt, welche bereits einen Suchdienst umsetzt, der auf den Grundlagen der sozialen Filterung basiert. Der zu entwickelnde Algorithmus basiert ebenfalls auf Grundlagen, Erfahrungen und Daten von Fooxx, so dass ein Überblick über diese Suchlösung hilfreich für die Entwicklung des Algorithmus ist.

3 Beschreibung des Suchdienstes „Fooxx“

Die Firma Global Brain Network GmbH, Fulda, entwickelt seit Anfang 2000 eine Suchmaschine mit den Namen „Fooxx“[7]. Diese Suchmaschine berücksichtigt bei einer Suchanfrage das Profil des Suchenden. Das zum Patent angemeldete Verfahren der personalisierten Suche wird wie folgt beschrieben: „Aufgabe der vorliegenden Erfindung ist es, ein Verfahren und eine Vorrichtung zur Ermittlung von relevanten Objekten für eine Suchanfrage eines Benutzers zu schaffen, bei denen Informationen über [alle] Benutzer, die auf einzelne Objekte bereits zugegriffen haben, für die Ermittlung der Relevanz der Objekte für spätere Suchanfragen berücksichtigt werden.“ [3,S.2]

Das Konzept basiert dabei auf den Prinzipien der in 2.7.5.2 vorgestellten sozialen Filterung, indem bei einer Suchanfrage eines Nutzers Informationen über andere Nutzer des Systems berücksichtigt werden.

Im Folgenden wird ein Überblick über die bestehende Implementierung dieses Verfahrens von Fooxx gegeben. Der im Rahmen dieser Diplomarbeit zu entwickelnde Algorithmus wird auf Daten und Programmteile des bestehenden Fooxx zugreifen. Es ist geplant, den zu entwickelnden Algorithmus später in das bestehende System mit einzubinden. Hierfür werden besonders die dazu in Bezug stehenden Teile der Suchmaschine Fooxx in diesem Kapitel hervorgehoben.

3.1 Client-/Serverseite von Fooxx im Überblick

Die bestehende Anwendung Fooxx besteht aus zwei Teilen – einer Client- und einer Server-Seite. Auf Client-Seite befindet sich eine im *Internet Explorer* integrierte Toolbar, die Informationen über den Nutzer an den Server übermittelt. Diese Informationen werden verarbeitet und entsprechend gespeichert.

Auf der Server-Seite befindet sich ein Webserver mit einer daran angebundenen Datenbank. Die übermittelten Daten der Toolbar sowie die Nutzerdaten werden in dieser Datenbank gespeichert. Stellt ein Nutzer des Systems eine Suchanfrage an den Server, wird zunächst eine Meta-Suchanfrage an drei verschiedene Suchdienste

[7] http://www.fooxx.com

(Google, AlltheWeb[8], Altavista[9]) gestellt. Diese Ergebnismenge wird zunächst nach dem Verfahren einer Metasuchmaschine sortiert. Dann werden gesammelte Interaktionsdaten von Nutzern der einzelnen Dokumente ausgewertet. Weiterhin wird anhand des Profils des suchenden Nutzers sowie vorhandenen Informationen anderer Nutzer die Trefferliste entsprechend neu sortiert.

Im den folgenden Abschnitten wird nun auf die Fooxx-Toolbar sowie auf die übermittelten Daten der Toolbar an den Fooxx-Server näher eingegangen. Anschließend wird die Erstellung eines Profils eines Fooxx-Nutzers erläutert und die einzelnen Such- und Sortierungsschritte auf Server-Seite beschrieben.

3.2 Fooxx-Toolbar

Auf der Client-Seite existiert die sog. Fooxx-Toolbar. Dieses Programm stellt eine Erweiterung des Browsers *Internet Explorer* dar, indem es sich in die bestehende Oberfläche des Browsers integriert und zusätzliche Funktionalität zur Verfügung stellt. (Abb. 2)

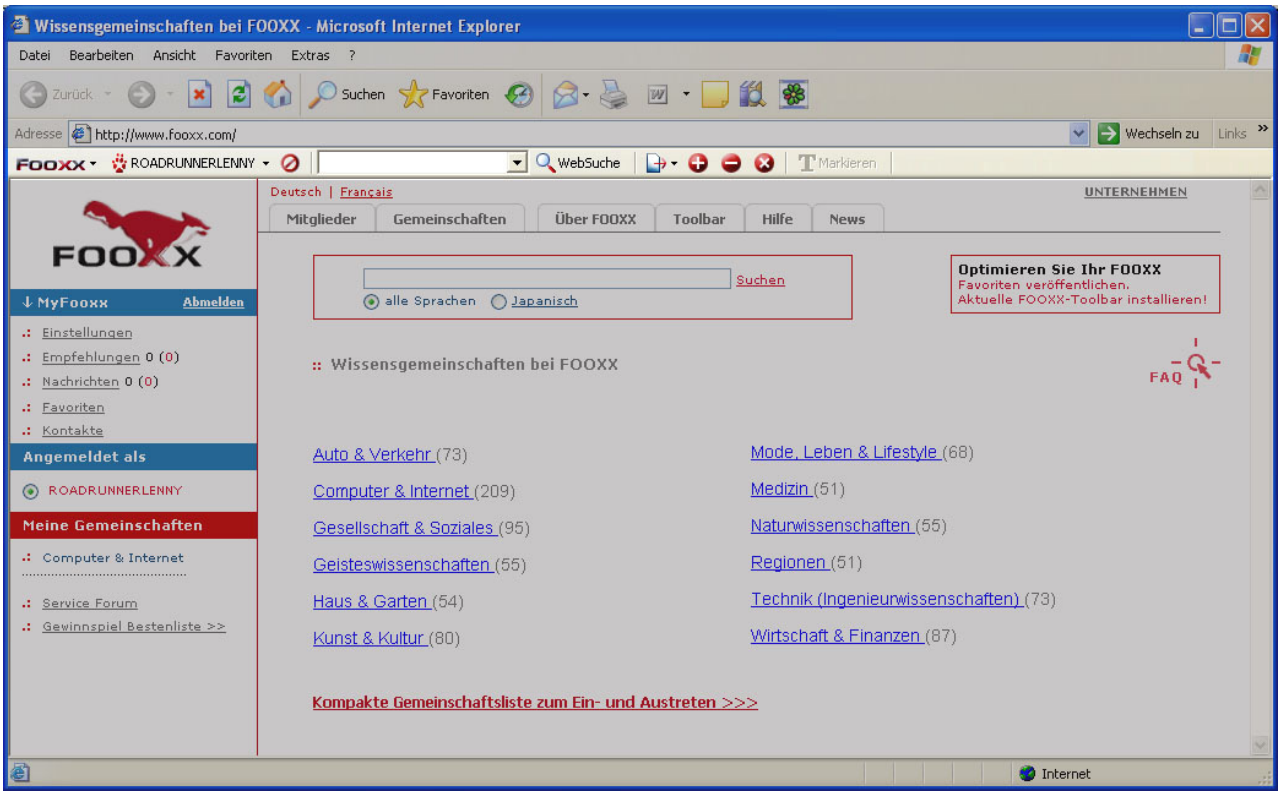


Abbildung 2: Screenshot der im Internet verfügbaren Fooxx-Toolbar (nicht abgedunkelter Bereich)

[8] http://www.alltheweb.de
[9] http://www.altavista.de

Damit die Toolbar wie gewünscht funktionieren kann, muss der Anwender sich bei Fooxx angemeldet haben. Nach der Anmeldung merkt die Toolbar sich die Nutzerdaten des Anwenders und beobachtet diesen bei seinen Aktivitäten im Internet. Dabei protokolliert die Toolbar jedes aufgerufene Dokument (in der Regel sind dies Webseiten) des Nutzers und errechnet aus der Dauer und Mausaktivität während der Betrachtung des Dokumentes einen Interaktionswert. Zudem gibt die Toolbar dem Anwender die Möglichkeit, ein momentan geöffnetes Dokument über einen Knopf in der Symbolleiste positiv oder negativ zu bewerten oder zu blockieren. Mit einer positiven Bewertung kann der Nutzer ein Dokument als besonders hilfreich markieren – analog mit einer negativen als nicht hilfreich. Mit einer Blockierung teilt der Nutzer dem System mit, das er dieses Dokument als nicht nützlich ansieht und es bei zukünftigen Suchanfragen nicht in der Ergebnisliste haben möchte.

Die so gesammelten Interaktionsdaten über die Aktivität während des Betrachten des Dokumentes sowie eine evtl. positive oder negative Bewertung oder Blockierung werden nach dem Verlassen des Dokumentes an den zentralen Server von Fooxx übermittelt und dort ausgewertet.

3.2.1 Auswertung der Interaktionsdaten

Zur Auswertung der Interaktionsdaten existiert eine Interaktionsmatrix, welche alle Interaktionen eines Dokumentes einem Nutzer zuordnet. Die Anzahl und Dauer der Interaktionen sowie evtl. eingegangene positive und negative Bewertungen und Blockierungen eines Dokumentes ergeben einen Wert, welcher als so genannter „innerer“ Wert bezeichnet wird. Dieser Wert macht eine Aussage über die Qualität des Dokumentes – je höher dieser Wert, umso nützlicher war dieses Dokument für den Nutzer. Ein Dokument, welches ein Nutzer lange aktiv betrachtet hat (Mausaktivitäten während der Betrachtung sichern auch die tatsächliche Anwesenheit eines Nutzers) kann als hilfreiches Dokument für den Nutzer angenommen werden. Ein nur kurz betrachtetes Dokument wird als nicht so hilfreich eingeordnet. Somit gibt es in der Interaktionsmatrix eine Zuordnung von Nutzern zu ihren betrachteten Dokumenten sowie zu den „inneren Werten“ dieses Dokumentes.

Der „innere Wert“ eines Dokumentes kann einmal als absoluter Wert berechnet werden, in dem alle Aktivitäten, Bewertungen und Blockierungen aller Nutzer

berücksichtigt werden. Der so erhaltene Wert ist dabei nicht spezifisch für einen Nutzer, sondern gibt eine Aussage über die Nutzung und Bewertung des Dokumentes aller Nutzer des Systems, die dieses Dokument aufgerufen haben. Zudem kann der innere Wert spezifisch für einen bestimmten Nutzer berechnet werden. Dazu werden für die Berechnung des Wertes nur die Aktivitäten, Bewertungen und Blockierungen der Nutzer berücksichtigt, die ein ähnliches Profil besitzen. Weiterhin besteht noch die Möglichkeit, den inneren Wert für eine Gruppe von Nutzern zu berechnen. Dabei werden nur die gesammelten Daten der Nutzer einbezogen, die einer bestimmten Gruppe von Nutzern zusammengefasst sind.

Bei jeder Berechnung des „inneren Wertes“ wird immer das Alter der gesammelten Interaktionsdaten über ein Dokument mit einbezogen. Je älter diese Daten sind, umso mehr wird der innere Wert des Dokumentes abgestuft. Dies bewirkt, dass evtl. veraltete und somit nicht mehr besuchte Dokumente mit der Zeit in Ihrem Wert abgestuft werden.

3.3 Arbeitsweise von Fooxx

Im Folgenden wird erläutert, wie ein Nutzer sich bei Fooxx registriert, bei Gemeinschaften von Fooxx anmeldet und auf diese Weise das Profil eines Nutzers erstellt wird. Anschließend wird der Ablauf einer Suchabfrage und die Abarbeitung dieser Suchabfrage von Fooxx als Metasuchmaschine dargestellt. Es wird erläutert, wie die abgefragten Sucherergebnisse anhand der ausgewerteten Interaktionsdaten bewertet werden und wie die Personalisierung der Suchergebnisse auf Basis der Nutzerprofile abläuft.

3.3.1 Registrierung und Profilerstellung

Für die Nutzung der vollen Funktionalität der personalisierten Suche von Fooxx ist es erforderlich, sich zunächst als Nutzer zu registrieren. Diese Registrierung erfolgt über das webbasierte Anmeldeformular von Fooxx. Beim Anmeldeprozess werden neben dem Nutzernamen, Passwort sowie persönlichen Daten dem Nutzer auch Gemeinschaften vorgestellt, denen er beitreten kann. Diese Gemeinschaften bestehen aus einer hierarchischen Liste von 12 Haupt-Gemeinschaften (z.B.: „Computer & Internet“) sowie einer großen Anzahl an Unter-Gemeinschaften (z.B. „Hardware“ als Untergemeinschaft von „Computer & Internet“). Durch Beitritt in eine dieser Gemeinschaften und der Angabe der persönlichen Kompetenz in diesen

Bereich („Zuhörer“, “Anwender“ oder „Experte“) wird dem Nutzer diese Gemeinschaft als Aspekt seines Profils mit der entsprechenden Gewichtung zugeordnet. Der Nutzer hat dabei die Möglichkeit, eigene Gemeinschaften anzulegen. Nach Abschluss des Anmeldeprozesses wird auf dem Server ein neuer Nutzer mit den angegebenen Aspekten seines erstellten Profils angelegt.

3.3.2 Suche von Fooxx

Stellt der Nutzer eine Suchanfrage an Fooxx, so wird zunächst die Menge an relevanten Dokumenten ermittelt. Dies wird zurzeit als Metasuche durch die Abfrage der Treffer anderer Suchmaschinen (Google, Altavista und AllTheWeb) umgesetzt. Die Treffer der anderen Suchmaschinen ergeben die Basis der Dokumentenmenge, die unter Berücksichtigung des Profils des Nutzers und den vorhandenen Interaktionsdaten und Profilen anderer Nutzer zu sortieren ist.

Es ergeben sich für die Sortierung der abgefragten Dokumentenmenge folgende Kriterien, die in drei Schritte zerlegt werden können.

1. Rang des Dokumentes in den abgefragten Suchmaschinen (Duplikate werden danach entfernt)
2. Innerer Wert des Dokumentes, absolut oder spezifisch zu einem Nutzer (in der Regel dem Suchenden Nutzer) oder einer Gemeinschaft
3. Profil des Suchenden im Verhältnis zu den Profilen der Nutzer der Dokumente

Die erhaltene Dokumentenmenge wird zunächst auf Duplikate überprüft. Für die Sortierung der Dokumentenmenge wird der bestehende Rang der Dokumente im ersten Sortierungsschritt mit einbezogen. Dies entspricht dem Sortierungsverhalten einer Metasuchmaschine. Aufgrund der im Theorieteil besprochenen Einschränkung einer Metasuchmaschine ist dabei die Abfrage von mehreren Suchmaschinen recht zeitintensiv. Da auch die interne Sortierung des inneren Wertes ebenfalls ein recht zeitintensiver Prozess ist, wird zugunsten einer schnelleren Antwortzeit die abzufragende Dokumentenmenge von den fremden Suchmaschinen recht klein gehalten. Bei der aktuellen Version von Fooxx werden bei der Erstellung der ersten Trefferseite von drei Suchmaschinen nur jeweils 10 Treffer abgefragt. Wünscht der Nutzer mehr Treffer zu sehen, so werden jeweils 10 weitere Treffer von jeder Suchmaschine abgefragt, usw. Eine Abfrage und Auswertung von z.B. den ersten

100 Treffern einer Suchmaschine würde eine recht lange Antwortzeit zur Folge haben (ca. 30sec.) und ist für den normalen Nutzer in der Regel nicht akzeptabel.

Anschließend wird der suchende Nutzer betrachtet. Die weitergehende Sortierung ermittelt den inneren Wert eines jedes Dokumentes spezifisch zum suchenden Nutzer. (Auf Wunsch des Suchenden kann die Ermittlung des inneren Wertes auch spezifisch zu einer vom Nutzer angegeben Gemeinschaft erfolgen). Ist der suchende Nutzer nicht bekannt (z.B. weil ein nicht registrierter Nutzer die Suchabfrage stellt), wird der absolute innere Wert jedes Dokumentes einbezogen. Unter Berücksichtigung der vorherigen Sortierung (dem Rang der abgefragten Suchmaschinen) wird die Ergebnisliste nach den inneren Werten neu sortiert und dem Suchenden angezeigt.

Als letztes werden die Profile der Nutzer der Dokumente mit dem Profil des suchenden Nutzers verglichen. Je höher die Übereinstimmung mit dem suchenden Nutzer, umso besser wird die Relevanz des Dokumentes bewertet. Dabei basieren die Profile der Nutzer auf den Gemeinschaften, den diese Nutzer beigetreten sind (z.B. „Hardware“), sowie ihren Angaben zu ihrer Kompetenz in dieser Gemeinschaft (z.B. „Experte“). Ein Vergleich der Profile zweier Nutzer wird dabei auf die Übereinstimmung der Gemeinschaften zurückgeführt. Bei diesem Punkt der Sortierung der Dokumentenmenge, der die eigentliche Personalisierung von Fooxx darstellt, existieren Probleme, die im Folgenden näher erläutert werden. Da die Personalisierung von Fooxx mit Problemen behaftet ist, wird in Kapitel 4 ein Algorithmus entwickelt, der diese Probleme beheben soll.

3.4 Probleme bei der Personalisierung des aktuellen Fooxx

Das bestehende System der Gemeinschaften ist durch die permanenten Erweiterungen der Nutzer inzwischen sehr groß geworden. Es gibt eine unübersichtliche Anzahl von über 1000 bestehenden Gemeinschaften. Zudem ist auch der erhoffte Nutzen der Gemeinschaften als Aspekt eines Profils eines Nutzers nach den Erfahrungen der Fooxx-Mitarbeiter nur beschränkt gegeben.

Diese Probleme haben verschiedene Ursachen. Zum einem kann jeder Nutzer sein eigene Gemeinschaft anlegen, was zu einer großen Masse an Gemeinschaften führt.

Ein neuer Nutzer ist nun gezwungen, sich einen Überblick über diese große Anzahl an Gemeinschaften zu verschaffen, um sich in die für sein Profil relevanten Gemeinschaften einzutragen. Dies kann dazu führen, dass der Nutzer nicht die Zeit und Lust hat, sich mit dem Vorgang der Profilerstellung länger auseinanderzusetzen und diesen somit vorzeitig abbricht. Ein weiteres Problem stellt die Aktualisierung der Profile dar: Da auch dies durch den Nutzer zu geschehen hat und ein zeitaufwändiger Prozess ist, wird er es nicht sehr häufig durchführen. Zudem stellt die Vergleichbarkeit der Profile ein großes Problem dar: Je größer die Anzahl der Nutzer, mit denen ein Profil verglichen wird, umso geringer wird die Aussage dieses Vergleiches. Eine Erklärung dafür ist, dass eine Übereinstimmung in den Hauptkategorien (wie z.B. „Computer & Internet“) leicht erreicht wird, in den Unterkategorien (wie z.B. „Hardware“) eher seltener. Vergleich man ein Profil nun mit mehreren Profilen, so wird entweder nur eine allgemeine Übereinstimmung in den Hauptkategorien oder nur eine sehr niedrige Übereinstimmung in den Unterkategorien ausgemacht werden können. Da die Kategorien vom Nutzer teilweise unvollständig eingepflegt und aktualisiert werden, sind verlässlich Aussagen über die tatsächlichen Ähnlichkeiten der Nutzer untereinander schwer möglich.

Diese Probleme führen dazu, dass die Profilerstellung und Personalisierung von Fooxx verbesserungswürdig ist. Im Rahmen dieser Arbeit wird im Folgenden Kapitel eine personalisierte Relevanzbewertung entwickelt, die die bestehende Profilerstellung und Personalisierung von Fooxx ersetzen soll. Die bestehende Metasuchmaschine sowie die Bewertung der abgefragten Treffermenge von Fooxx mit Hilfe des inneren Wertes werden dabei unverändert beibehalten. Die Profilerstellung und die Relevanzbewertung der Dokumente aufgrund des Vergleiches der Nutzerprofile werden neu entwickelt.

4 Algorithmus zur personalisierten Relevanzbewertung

In diesem Kapitel wird ein Algorithmus entwickelt, mit dem die gefundenen relevanten Dokumente eines nicht personalisierten Suchverfahrens nach der personalisierten Relevanz für den suchenden Nutzer neu bewertet werden. Die Grundlage dieser Relevanzbewertung ist das Profil des Suchenden sowie die Profile der Nutzer der Dokumente. Anhand der erstellten Bewertung kann eine personalisierte Sortierung der Dokumentenmenge erfolgen, welche später im bestehenden Fooxx die Personalisierung der Dokumente und die Profilerstellung ersetzen kann. Das Prinzip der Ermittlung des inneren Wertes sowie die grundlegende Metasuchfunktion in Fooxx bleiben dabei bestehen.

4.1 Konzept der Relevanzbewertung und Profilerstellung

Es gibt eine Anzahl von Nutzern und eine Menge von Dokumenten im System. Jeder Nutzer von Fooxx besitzt ein persönliches Profil. Das Profil eines jeden Nutzers wird aus der Anzahl der benutzten Dokumente erstellt. Eine Ähnlichkeit der Nutzer ist anhand eines Vergleiches der verwendeten Dokumente festzumachen. Je höher die Anzahl an gleichen verwendeten Dokumenten, umso höher die Ähnlichkeit der Nutzer. Wird nun eine Suchanfrage an das System gestellt, so wird aus den Dokumenten eine zunächst unsortierte Menge an relevanten Dokumenten ermittelt. Diese Dokumente werden dann folgendermaßen sortiert: Wurde ein Dokument bereits von vielen Nutzern benutzt, die eine hohe Ähnlichkeit mit dem bestehenden Nutzer aufweisen, so wird es in der Sortierung weiter oben angezeigt. Dokumente, die von Nutzern mit einer geringeren Ähnlichkeit benutzt wurden, werden niedriger bewertet und weiter unten einsortiert.

Nach diesen Kriterien ergibt sich nun eine neue Sortierung der Auswahl, die dem Nutzer zuerst Dokumente präsentiert, die eine hohe Qualität besitzen und auch von ähnlichen Nutzern (also von Nutzern mit wahrscheinlich gleichen Interessen) benutzt wurden.

Die im Rahmen dieser Arbeit zu entwickelnde Relevanzbewertung soll in das bestehende Fooxx später implementiert werden. Dabei soll diese Relevanzbewertung in Fooxx die Personalisierung der Dokumente ersetzen. Unberührt davon bleibt die Abfrage anderer Suchmaschinen, die erste Sortierung der

einzelnen Dokumente nach dem jeweiligen Rang in den abgefragten Suchmaschine und die zweite Sortierung nach dem inneren Wert der Dokumente spezifisch zum Profil des Nutzers. Die im Folgenden vorgestellte Relevanzbewertung ermittelt einen beschreibenden Wert für ein Dokument, und definiert dabei implizit das Profil eines Nutzers. Das Konzept dieser Relevanzbewertung sowie die (implizite) Profilerstellung werden im Folgenden detaillierter erläutert.

Die Erstellung des entsprechenden Algorithmus zur Relevanzbewertung wird in dem nächsten Kapitel eingehender erläutert. Bevor der Algorithmus im Anschluss an dieses Kapitel in das bestehende Fooxx integriert wird, wird er in einer entwickelten Testumgebung mit bestehenden Daten von Fooxx einem ersten Test unterzogen und auf seinen Zweck hin überprüft.

4.2 Definition von Nutzern, Dokumenten und ihre implizite Profilerstellung

Es gibt eine Menge von bekannten Nutzern und eine Menge von bekannten Dokumenten. Es sei N die Menge aller Nutzer mit

$$N = \{n_1, n_2, ..., n_k\} \tag{F.1}$$

und D die Menge aller Dokumente mit

$$D = \{d_1, d_2, ..., d_l\}\,. \tag{F.2}$$

Durch die Benutzung eines Dokumentes durch den Nutzer (z.B. durch das Betrachten eines Dokumentes) wird dem Nutzer das entsprechende Dokument zugeordnet. Weiterhin wird auch dem Dokument dieser Nutzer zugeordnet. Zu jedem Nutzer gehört somit eine Menge aller von ihm benutzten Dokumente, und zu jedem Dokument gehört eine Menge aller Nutzer, die dieses Dokument benutzt haben.

Die Menge aller von dem i-ten Nutzer benutzten Dokumente sei gegeben durch die Menge Φ_i. Diese Menge ist definiert durch:

$$\forall i \in \overline{1,k} : \Phi_i = \{d_{j_{i,1}}, d_{j_{i,2}}, ..., d_{j_{i,p_i}}\} \subset D \text{ , wobei } j_{i,x} \in \overline{1,l}, x \in \overline{1,p_i} \tag{F. 3}$$

x stellt dabei einen durchnummerierten Index von 1 bis p_i dieser Menge dar. Die Menge Φ_i definiert sich als Teilmenge der Dokumentenmenge D, da ein Nutzer

durch die Menge seiner benutzten Dokumente d_j beschrieben werden kann. Diese Teilmenge kann als „Profil des Nutzers“ verstanden werden.

Die Menge aller Nutzer, die ein bestimmtes Dokument benutzt haben, sei gegeben durch die Menge Ψ_j. Diese Menge ist definiert durch:

$$\forall j \in \overline{1,l}: \ \Psi_j = \{n_{i_{j,1}}, n_{i_{j,2}}, ..., n_{i_{j,q_j}}\} \subset N \text{ , wobei } i_{j,y} \in \overline{1,k}, y \in \overline{1,q_j} \qquad \text{(F. 4)}$$

y stellt dabei einen durchnummerierten Index von 1 bis q_j dieser Menge dar. Die Menge Ψ_j definiert sich als Teilmenge der Nutzermenge N. Sie beschreibt alle Nutzer n_i, die ein bestimmtes Dokument benutzt haben. Sie kann als „Profil des Dokumentes“ verstanden werden. Zur Verdeutlichung ein Beispiel:

Beispiel 1

Es werden zwei Elemente n_{r_1}, n_{r_2} ($r_1, r_2 \in \overline{1,k}$) aus der Menge der Nutzer N sowie vier Elemente $d_{s_1}, d_{s_2}, d_{s_3}, d_{s_4}$ ($s_1, ..., s_4 \in \overline{1,l}$) aus der Menge der Dokumente D betrachtet. Ein Nutzer n_{r_1} hat die drei Dokumente $d_{s_1}, d_{s_2}, d_{s_3}$ benutzt. Ein zweiter Nutzer n_{r_2} benutzt zwei gleiche Dokumente d_{s_1}, d_{s_3} und ein anderes Dokument d_{s_4}. Es ergibt sich folgende schematische Abbildung zur Darstellung der Beziehung zwischen Dokumenten und Nutzer:

$$\begin{array}{c} \\ d_{s_1} \\ d_{s_2} \\ d_{s_3} \\ d_{s_4} \end{array} \begin{array}{c} \begin{array}{cc} n_{r_1} & n_{r_1} \end{array} \\ \begin{pmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{1} \\ 0 & \mathbf{1} \end{pmatrix} \end{array}$$

1: der Nutzer hat das Dokument benutzt
0: das Nutzer hat das Dokument nicht benutzt

Es ergeben zudem folgende Profile für die Nutzer und die Dokumente.

$\Phi_{r_1} \supset \{d_{s_1}, d_{s_2}, d_{s_3}\}, \Phi_{r_2} \supset \{d_{s_1}, d_{s_3}, d_{s_4}\}$ als Profile der Nutzer und

$\Psi_{s_1} \supset \{n_{r_1}, n_{r_2}\}, \Psi_{s_2} \supset \{n_{r_1}\}, \Psi_{s_3} \supset \{n_{r_1}, n_{r_2}\}, \Psi_{s_4} \supset \{n_{r_2}\}$ als Profile der Dokumente.

Man kann sehen, dass das Nutzerprofil Φ_{r_1} des Nutzers n_{r_1} durch die Dokumente $d_{s_1}, d_{s_2}, d_{s_3}$ ergänzt wurde. Zudem wurden die Dokumentenprofile $\Psi_{s_1}, \Psi_{s_2}, \Psi_{s_3}$ um den Nutzer n_{r_1} ergänzt. Das Nutzerprofil Φ_{r_2} des Nutzers n_{r_2} wurde mit den Dokumenten $d_{s_1}, d_{s_3}, d_{s_4}$ ergänzt, sowie die Dokumentenprofile $\Psi_{s_1}, \Psi_{s_3}, \Psi_{s_4}$ um den Nutzer n_{r_2} .

4.3 Ähnlichkeitswerte von Dokumenten

Es kann ein Nutzerprofil Φ_i mit dem Profil eines anderen Nutzers Φ_j verglichen werden, um eine Überstimmung beider Profile zu ermitteln. Dazu wird die Anzahl der gleichen Dokumente beider Profile in das Verhältnis zur gesamten Anzahl an Dokumenten beider Profile gesetzt. Das resultierende Ergebnis ist ein Wert, der die Ähnlichkeit zwischen zwei Nutzerprofilen beschreibt. Dieser Wert wird als Ähnlichkeitswert a_{ij} bezeichnet. Für den Ähnlichkeitswert gilt:

$$a_{ij} = \frac{|\Phi_i \cap \Phi_j|}{|\Phi_i \cup \Phi_j|} \tag{F.5}$$

wobei $|\cdot|$ als Funktion zur Bestimmung der gesamten Anzahl an Elementen in einer Menge definiert ist. Der Wertebereich der Funktion $|\cdot|$ ist die Menge aller nicht negativen Ganzzahlen mit $N_0 = \{0,1,2,...\}$. Des weiteren gilt, dass die Schnittmenge von Φ_i und Φ_j eine Untermenge der Vereinigungsmenge beider Mengen ist. Da diese Schnittmenge immer kleiner oder gleich der Vereinigungsmenge ist, gilt

$$\Phi_i \cap \Phi_j \subseteq \Phi_i \cup \Phi_j \Rightarrow |\Phi_i \cap \Phi_j| \leq |\Phi_i \cup \Phi_j|, \tag{F.6}$$

woraus folgt, dass $0 \leq a_{ij} \leq 1$ und der Ähnlichkeitswert a_{ij} somit immer in den Grenzen $[0,1]$ liegt.

Analog kann auch ein Ähnlichkeitswert für die Übereinstimmung der Nutzer von Dokumenten gebildet werden, wird aber für die weiterführende Ausarbeitung nicht benötigt.

4.4 Grundlagen des Relevanzbewertungsalgorithmus

Entsteht eine Suchanfrage eines Nutzers (des sog. „Suchenden“) zur Auffindung von Informationen in den dafür relevanten Dokumenten aus der Dokumentenmenge, wird zunächst die Menge an relevanten Dokumenten R aus der Gesamtmenge der Dokumente über ein nicht personalisiertes Suchverfahren ausgesucht. Bei dem Suchverfahren kann es sich z.B. um eine klassische Volltextsuche handeln, die alle Dokumente findet, die die übergebenen Schlüsselwörter einer gestellten Suchabfrage beinhalten. Die gefundenen Dokumente werden dann nach der Häufigkeit des Vorkommens der Schlüsselwörter in dem jeweiligen Dokument sortiert. Wichtig ist, dass das Suchverfahren eine Menge an relevanten Dokumenten findet, und diese Menge sortiert ist, ohne das persönliche Profil des Suchenden zu berücksichtigen. Für die zahlenmäßige Bewertung zur Qualität eines Suchverfahrens kann hierbei auf die Werte der Abdeckung *(Recall)* und der Präzision *(Precision)* zurückgegriffen werden. (Vergleiche Kapitel 2.5.)

Die Menge der gefundenen relevanten Dokumente R ist somit nach einem nicht personalisierten Verfahren sortiert. Um die Elemente von R nach einem personalisierten Verfahren neu zu ordnen, ist es notwendig, den Dokumenten einen beschreibenden vergleichbaren Wert zuzuordnen. Dieser beschreibende Wert kann als Gewicht des Dokumentes verstanden werden. Die einzelnen Elemente können dann mit Hilfe der bekannten arithmetischen Ordnungsrelation $\geq$ verglichen und sortiert werden.

Es ist nun eine Funktion g zu finden, die das Ermitteln des beschreibenden Wertes (des Rangs) eines Dokumentes vornimmt. Grundlage für diese Funktion soll das Profil des Suchenden sowie die Profile der Nutzer der gefundenen relevanten Dokumente sein. Die Funktion vergleicht dazu die Ähnlichkeitswerte zwischen dem Suchenden und den Nutzern des Dokumentes und errechnet daraus den beschreibenden Wert jedes Dokumentes. Dabei geht die Funktion vom Folgenden aus: Dokumente, die von ähnlichen Nutzern verwendet wurden, sind von höherer Relevanz für den Suchenden. Sind alle beschreibende Werte ermittelt, so können die Dokumente sortiert werden. Das Dokument mit dem höchsten beschreibenden Wert ist wahrscheinlich das wichtigste bzw. relevanteste Dokument für den Suchenden, da

es von einer höheren Anzahl von ähnlicheren Nutzern benutzt wurde als andere Dokumente.

4.5 Abbildung der Bewertungsfunktion

Es gilt, dass jedem Dokument d_s eine Anzahl von Nutzern n_i zugeordnet ist. Zudem kann für jeden Nutzer der Ähnlichkeitswert a_{ij} des Nutzers n_i im Verhältnis mit dem Suchenden (dem Nutzer n_j) ermittelt werden. Die Funktion g hat als Argumente eine endliche Menge der Ähnlichkeitswerte a_{ij} der Nutzer n_i des Dokumentes im Vergleich zu dem suchenden Nutzer n_j. Es werden dabei nur die Ähnlichkeitswerte mit einem Wert größer 0 berücksichtigt, da für die Ermittlung des beschreibenden Wertes eines Dokumentes die Nutzer mit keiner Ähnlichkeit zum Suchenden bei dem Verfahren nicht mit einbezogen werden sollen. Das Ergebnis der Funktion ist ein nicht negativer endlicher Wert (einschließlich 0).

Es gilt somit für die Abbildung der Funktion $g : A \rightarrow \mathbf{R}^+ \cup \{0\}$, wobei die Menge A die Menge aller endlichen Mengen der Zahlen vom Range (0,1] darstellt und es gezeigt werden kann, dass A die Menge aller Ähnlichkeitswerte darstellt. Diese Menge wird auf die Menge der reellen Zahlen $\mathbf{R}^+$ inklusive der 0 abgebildet. Wenn vom Nutzer n_i eine Suchabfrage erfolgt, wird die Ähnlichkeit des i-ten Nutzers zum Suchenden bezüglich des j-ten Dokumentes als A_{ij} bezeichnet und definiert als:

$$\forall i \in \overline{1,k} : \forall j \in \overline{1,l} : A_{i,j} = \{a_{i,k} \mid k \in \{k \mid n_k \in \Psi_j\} \,\&\, a_{i,k} > 0\} \tag{F.7}$$

Da die Menge $A_{i,j}$ aus einer endlichen Menge der Zahlen vom Range (0, 1] besteht und die Menge A ebenfalls eine Menge der Zahlen desselben Ranges ist, gilt:

$$\forall i \in \overline{1,k} : \forall j \in \overline{1,l} : A_{i,j} \in A \tag{F.8}$$

Es kann somit gefolgert werden, dass die Menge A die Menge aller möglichen Ähnlichkeitswerte abdeckt. Bei der vollständigen Betrachtung der Ähnlichkeitswerte ist einzuschränken, dass der suchende Nutzer nicht mit sich selbst verglichen werden soll.

Nachdem die Abbildung der Funktion g charakterisiert wurde, können nun Anforderungen definiert werden, um der Funktion einen sinnvollen Rahmen zu geben.

4.6 Anforderungen an die Bewertungsfunktion

Folgende Bedingungen werden an die Funktion g gestellt, damit diese einen beschreibenden Wert zur Aussage über die Wichtigkeit eines Dokumentes für den Suchenden ermitteln kann:

B1: Steigt die Anzahl der Ähnlichkeitswerte, die einem Dokument zugeordnet sind, so steigt auch das Ergebnis der Funktion g.Das heißt, die Funktion soll monoton mit der Anzahl der Nutzer, die das Dokument genutzt haben, wachsen. Je mehr Nutzer, desto größer auch das Ergebnis.

Es gilt:

$$g(\{a_{1,j}, a_{2,j}, ..., a_{h,j}\}) < g(\{a_{1,j}, a_{2,j}, ..., a_{h,j}, a_{(h+1),j}\}) \text{, wobei } \forall i = \overline{1, h+1} : a_{i,j} > 0 \qquad \text{(Bed. 1)}$$

B2: Für alle Ähnlichkeitswerte a_{ij} gilt: steigt der Ähnlichkeitswert selbst, so steigt auch das Ergebnis der Funktion g. Die Funktion soll monoton mit den Ähnlichkeitswerten wachsen. Je ähnlicher die Nutzer, desto größer auch das Ergebnis.

Es gilt:

$$g(\{a_{1,j}, a_{2,j}, ..., a_{u,j}, ..., a_{h,j}\}) < g(\{a_{1,j}, a_{2,j}, ..., a_{u,j} + c, ..., a_{h,j}\}), \qquad \text{(Bed.2)}$$

wobei *c>0* und $\forall i = \overline{1, h} : a_{i,j} > 0$

B3: Das Ergebnis der Funktion g einer leeren Menge von Ähnlichkeitswerten ist 0. Falls keine ähnlichen Nutzer das Dokument benutzt haben, so muss das Ergebnis 0 sein, sollte es aber mindestens einen ähnlichen Nutzer geben, so muss das Ergebnis größer als 0 sein.

Es gilt:

$$g(X) = 0 \Leftrightarrow X = \varnothing \qquad \text{(Bed.3)}$$

B4: Um das Ergebnis der Funktion nachher sinnvoll abbilden zu können (z.B. als Prozentzahl), muss es von g beschränkt werden. Es wird angenommen, dass es kleiner als 1 sei.

Es gilt:

$$\forall X \in F : g(X) < 1 \qquad \text{(Bed.4)}$$

Diese vier Bedingungen geben den Rahmen wieder, der an eine Funktion zur Ermittlung eines beschreibenden Wertes für die Wichtigkeit eines Dokumentes für den Suchenden gestellt werden muss. Im Folgenden wird eine Funktion vorgestellt, die diese Anforderungen erfüllt.

4.7 Vorgeschlagene Bewertungsfunktion g_1

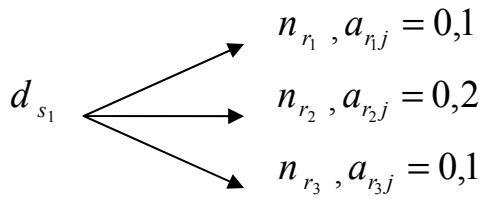
Es wird davon ausgegangen, dass die Summe aller Ähnlichkeitswerte der einzelnen Nutzer n_i im Verhältnis zum Suchenden eine Aussage über die Wichtigkeit des Dokumentes für den suchenden Nutzer n_j machen kann. Dies kann beschrieben werden durch:

$$\sum_{i=1}^{n} a_{ij} \geq 0 \qquad \text{(F.9)}$$

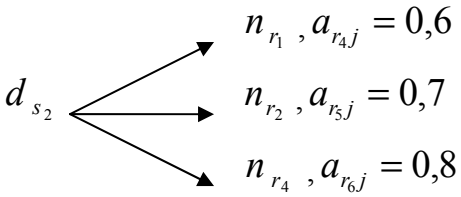
wobei j sich auf den Suchenden Nutzer n_j bezieht und ein konstanter Wert ist. Da a_{ij} immer in den Grenzen (0,1] liegt, ist stets ein nicht negatives Ergebnis zu erwarten (da n auch 0 sein kann, ist auch der Wert 0 nicht auszuschließen). Hierzu ein Beispiel.

Beispiel 2:

Gegeben seien die Ähnlichkeitswerte $a_{r_1 j} = 0{,}1$, $a_{r_2 j} = 0{,}2$ und $a_{r_3 j} = 0{,}1$ von drei Nutzern n_{r_1}, n_{r_2}, n_{r_3} eines Dokumentes d_{s_1} im Bezug zum Nutzer n_j. Zudem seien die Ähnlichkeitswerte $a_{r_4 j} = 0{,}6$, $a_{r_5 j} = 0{,}7$ und $a_{r_6 j} = 0{,}8$ von zwei gleichen Nutzern n_{r_1}, n_{r_2} und einem anderen Nutzer n_{r_4} eines Dokumentes d_{s_2} im Bezug zum Nutzer n_j gegeben.



$$\sum_{i=1}^{3} a_{r_i j} = 0{,}1 + 0{,}2 + 0{,}1 = 0{,}4$$



$$\sum_{i=4}^{6} a_{r_i j} = 0{,}6 + 0{,}7 + 0{,}8 = 2{,}1$$

Die Summe der drei Ähnlichkeitswerte ergibt in diesem Beispiel für das Dokument d_{s_1} 0,4 und für das Dokument d_{s_2} 2,1. Das zweite Dokument mit dem höheren Wert weist eine höhere Anzahl an ähnlichen Nutzern auf und ist somit das wahrscheinlich wichtigere Dokument für den Suchenden Nutzer n_j.

Es kann allerdings bei einer großen Menge an Nutzern mit einer durchweg geringen Ähnlichkeit zum Suchenden vorkommen, dass das dazugehörige Dokument eine höhere Bewertung erreicht als ein Dokument mit einer kleinen Menge von Nutzern und einer sehr hohen Ähnlichkeit zum Suchenden. Dies kann durch folgendes Beispiel verdeutlicht werden:

Beispiel 3:

Gegeben seien zwei Dokumente d_{s_1} und d_{s_2}. Für das erste Dokument d_{s_1} gibt es 100 Nutzer $n_{i_{s_1,1}},\ldots,n_{i_{s_1,100}}$ mit den gleichen Ähnlichkeitswerten $a_{ji_{s_1,1}} = 0{,}1,\ldots..,a_{ji_{s_1,100}} = 0{,}1$ im Verhältnis zum Suchenden Nutzer n_j. Für das zweite Dokument d_{s_2} gibt es 10 Nutzer $n_{i_{s_2,101}},\ldots,n_{i_{s_2,111}}$ mit ebenfalls gleichen Ähnlichkeitswerten $a_{ji_{s_2,101}} = 0{,}9,\ldots,a_{ji_{s_2,111}} = 0{,}9$ im Verhältnis zum Suchenden n_j.

Als beschreibenden Wert für Dokument d_{s_1} erhält man:

$$\sum_{i=1}^{100} a_{ji_{s_1,i}} = 0,1 + ... + 0,1 = 0,1 \cdot 100 = 10$$

Als beschreibenden Wert für Dokument d_{s_2} erhält man:

$$\sum_{i=101}^{111} a_{ji_{s_2,i}} = 0,9 + ... + 0,9 = 0,9 \cdot 10 = 9$$

Man kann sehen, dass das Dokument d_{s_2} mit einer sehr hohen Anzahl von sehr ähnlichen Nutzern keinen so hohen beschreibenden Wert erhält wie ein Dokument mit einer großen Masse an Nutzern, deren Ähnlichkeitswert aber sehr klein ist. Je größer die Anzahl der Nutzer wird, umso geringer ist der Einfluss des Ähnlichkeitswertes auf die Sortierung.

Um den abnehmenden Einfluss des Ähnlichkeitswertes anpassen zu können, kann es sinnvoll sein, einen hohen Ähnlichkeitswert höher zu bewerten als einen Niedrigen. Dies kann durch Potenzieren jedes Ähnlichkeitswertes der Formel 9 mit einem konstanten Faktor k geschehen, was einen hohen Wert nur schwach und einen niedrigen Wert stark beeinflussen würde. (*Man beachte*: a_{ij} liegt in den Grenzen (0,1].) Damit das gewünschte Ergebnis erzielt wird, muss k größer gleich 0 sein. Hierbei ist zu beachten, das ein k im Bereich $0 \leq k \leq 1$ diesen Effekt nur schwach ausprägt bzw. für $k = 0$ die Ähnlichkeiten ganz unberücksichtig lässt und nur die Anzahl der Nutzer berücksichtigt. Welcher exakte Wert die sinnvollsten Ergebnisse liefert, ist später durch praktische Tests herauszufinden. Es kann sich evtl. auch herausstellen, dass eine schwächere Einflussnahme der Ähnlichkeitswerte unter bestimmten Umständen besser ist. Die Einführung eines variablen Faktors k zur Einflussnahme auf die Bedeutung der Ähnlichkeitswerte ist auf jeden Fall sinnvoll.

$$\sum_{i=1}^{n} a_{ij}{}^{k} \geq 0 \quad \text{mit } k \geq 0 \tag{F.10}$$

Die Aussage, dass eine große Anzahl von hohen Ähnlichkeitswerten der Nutzer im Verhältnis zum Suchenden für ein Dokument ein höheres Ergebnis zurückliefert, wird

durch Formel 10 nicht widersprochen und Bedingung 1 ist erfüllt. Da k größer gleich 0 ist, wird ebenfalls Bedingung 2 erfüllt, nämlich dass bei einem höheren Ähnlichkeitswert auch das Ergebnis der Funktion steigt. Werden keine Ähnlichkeitswerte übergeben, so ist das Ergebnis der Funktion 0, und werden Ähnlichkeitswerte übergeben, ist das Ergebnis größer 0. Somit ist auch Bedingung 3 erfüllt. Allerdings widerspricht Formel 10 der Bedingung 4, da das Ergebnis der Funktion größer als 1 sein kann.

Um Bedingung 4 gerecht zu werden, muss die Formel 10, deren Grenzen des Wertebereichs in dem Bereich $[0,\infty)$ liegen, so transformiert werden, dass die oben genannte Bedingung weiterhin erfüllt und der Wertebereich auf die Grenzen $[0,1)$ verschoben wird. Dazu betrachten wir folgende Hilfsfunktion f :

$$f(x) = 1 - \frac{1}{1+x} \tag{F.11}$$

Mit Formel 11 haben wir eine Funktion, die sich unter Berücksichtigung der Bedingungen 1, 2 und 3 im Wertebereich in den Grenzen $[0,1)$ befindet und dabei ausschließlich die 0 auf 0 abbildet und das monotone Wachstum von Bedingung 1 und Bedingung 2 unberührt lässt. Da der Grenzwert der Funktion mit $\lim(f(x))_{x\to\infty} = 1$ gegen 1 läuft, ist nun auch Bedingung 4 erfüllt. Mit Hilfe von Formel 11 kann nun durch Zusammenführen mit Formel 10 die Funktion g_1 beschrieben werden:

$$g_1(A) = 1 - \frac{1}{1+\sum_{i=1}^{n} a_{ij}{}^{k}}, \tag{F.12}$$

wobei $g_1(\cdot) \in [0,1)$, $k \geq 0$, $0 \leq a_{ij} \leq 1$.

Um Formel 12 zu verdeutlichen, ein Beispiel:

Beispiel 4

Es werden die gleichen Dokumente, Nutzer sowie Ähnlichkeitswerte zum Suchenden angenommen wie in Beispiel 3. Als Faktor k wird beispielhaft der Wert 2 ausgewählt.

Als beschreibenden Wert für Dokument d_{s_1} erhält man:

$$1-\frac{1}{1+\sum_{i=1}^{100} a_{ji_{s1,i}}{}^2} = 1-\frac{1}{1+(0{,}1^2+\ldots+0{,}1^2)} = 1-\frac{1}{1+(0{,}1^2*100)} = 0{,}5$$

Als beschreibenden Wert für Dokument d_{s_2} erhält man:

$$1-\frac{1}{1+\sum_{i=101}^{111} a_{ji_{s2,i}}{}^2} = 1-\frac{1}{1+(0{,}9^2+\ldots+0{,}9^2)} = 1-\frac{1}{1+(0{,}9^2*10)} \approx 0{,}89$$

Das Dokument d_{s_2} hat einen höheren beschreibenden Wert (nämlich 0,89) und kann somit als wichtiger für den Suchenden gewertet werden als Dokument d_{s_1} .

Die hier vorgestellte Funktion erfüllt die oben genannten vier Bedingungen und ist eine mögliche Funktion zur Ermittlung eines beschreibenden Wertes für die Relevanz eines Dokumentes für den Suchenden. Dabei werden die Profile des Suchenden sowie die Profile der Nutzer berücksichtigt, die dieses Dokument bereits genutzt haben. Anhand der zurück gelieferten Ergebnisse kann eine Menge von gefundenen relevanten Dokumenten personalisiert geordnet werden. Im weiteren Verlauf dieser Arbeit wird diese Formel daraufhin untersucht, ob sie in einer praktischen Umsetzung im Rahmen einer Testumgebung auch sinnvolle und vertretbare Ergebnisse liefert.

5 Implementierung von Bewertungsalgorithmus und Testumgebung

In diesem Kapitel wird erläutert, wie der beschriebene Relevanzbewertungs-Algorithmus zur Bestimmung eines beschreibenden Wertes eines Dokumentes in geeigneter Form implementiert und auf seine Aufgabenerfüllung hin getestet werden kann. Dazu ist zunächst ein Konzept für die Implementierung zu entwickeln. Weiterhin ist eine Testumgebung sowie ein Testablauf zu konzeptionieren, in der sinnvolle Testdaten bereitgestellt und die Ergebnisse des Algorithmus betrachtet, verglichen und ausgewertet werden können.

5.1 Anforderungsbeschreibung an die Implementierung

Für die praktische Implementierung ist eine möglichst genaue Beschreibung der Anforderungen zu finden.

Zunächst ist der beschriebene Bewertungsalgorithmus zu implementieren. Es gibt folgende Anforderungen an die Implementierung:

Funktional:

1. Anforderung (Anf. 1)
   Die Implementierung muss die in Kapitel 4.7 beschriebene Funktionalität der Funktion g_1 abbilden. Nach Übergabe einer Menge an Ähnlichkeitswerten muss das Programm daraus den beschreibenden Wert bilden können.

Nicht-funktional:

2. Anforderung (Anf. 2)
   Der Algorithmus muss in einem eigenen Package abgebildet sein. Damit wird sichergestellt, dass eine Portabilität des Algorithmus möglich ist.
3. Anforderung (Anf. 3)
   Eine Schnittstelle des Algorithmus ist erforderlich, die neben dem implementierten Algorithmus g_1 noch weitere Algorithmen zulässt.

Erläuterungen:

Zu Anforderung 1: Die Implementierung des Algorithmus soll folgende Anforderungen erfüllen: Eine Menge an Ähnlichkeitswerten werden übergeben. Die Implementierung ermittelt mit den übergebenen

Ähnlichkeitswerten einen beschreibenden Wert anhand der in 4.7 erklärten Formel g_1. Dieser Wert wird dann zur weiteren Verarbeitung zurückgeliefert.

Zu Anforderung 3 und 4: Das Auslagern des Algorithmus in ein eigenes Package stellt sicher, dass dieser nicht nur in der bestehenden Anwendung, sondern auch in weiteren Anwendungen verwendet werden kann. Durch das Definieren einer Schnittstelle können weitere Algorithmen entwickelt und implementiert werden. Da es Ziel ist, den Algorithmus nach der Implementierung und dem Testen in die bestehende Anwendung von Fooxx zu integrieren, sind diese Anforderungen sinnvoll.

Weiterhin ist eine Testumgebung zu entwickeln, mit der der entwickelte Relevanzbewertungs-Algorithmus getestet werden kann. Folgende Anforderungen werden an diese Testumgebung gestellt:

Funktional:

4. Anforderung (Anf. 4)
   Die Testumgebung muss eine Verbindung mit einer Datenbank herstellen können, um die erforderlichen Daten für den Test abrufen zu können.
5. Anforderung (Anf. 5)
   Die Testumgebung muss eine Verbindung mit verschiedenen Suchmaschinen zur Abfrage einer Trefferliste herstellen können.
6. Anforderung (Anf. 6)
   Die Testumgebung soll Messdaten liefern, die bei der Durchführung der Tests relevant sind.
7. Anforderung (Anf. 7)
   Die Testumgebung soll eine grafische Oberfläche bieten, mit denen Tests durchgeführt werden können.

Nicht-funktional:

8. Anforderung (Anf. 8)
   Die Testumgebung soll eine Schnittstelle bieten, über die eine unsortierte Dokumentenmenge bereitgestellt und eine sortierte Dokumentenmenge abgefragt werden kann.

Erläuterungen:

Zu Anforderung 4: Die Daten für die durchzuführenden Tests werden in einer bestehenden Datenbank bereitgestellt. Die Testumgebung muss in der Lage sein, auf diese Datenbank zuzugreifen. Die benutzte Datenbank wird dabei die bestehende Fooxx-Datenbank sein. Es soll aber auch möglich sein, durch entsprechende Modifikationen auf eine andere Datenbank zuzugreifen.

Zu Anforderung 5: Die Testumgebung soll Trefferlisten von Suchmaschinen abfragen. Diese abgefragten Dokumentenmengen dienen dazu, von dem Relevanzbewertungs-Algorithmus neu bewertet und anschließend sortiert zu werden.

Zu Anforderung 6: Die Testumgebung soll in der Lage sein, Messdaten über den Relevanzbewertungs-Algorithmus zu liefern. Diese Messdaten sollen hauptsächlich aus Zeitmessungen sowie detaillierten Informationen über die zu bewertenden Dokumente bestehen.

Zu Anforderung 7: Die Testumgebung soll eine grafische Oberfläche bereitstellen, über die der Nutzer die Testumgebung bedienen kann. Über diese Oberfläche können dann die unsortierten Dokumentenmengen der Suchmaschinen abgefragt, bewertet, sortiert und schließlich angezeigt werden. Ebenfalls werden hier die Messdaten zur Anzeige gebracht.

Zu Anforderung 8: In der Testumgebung soll eine Schnittstelle implementiert werden, über die eine unsortierte Dokumentenmenge bereitgestellt werden kann. Nach der Durchführung der Bewertung und Sortierung mit Hilfe des Relevanzbewertungs-Algorithmus der Dokumentenmenge soll diese über die Schnittstelle wieder abgefragt werden können. Der Vorteil dieser Schnittstelle besteht darin, dass die Testumgebung nicht ausschließlich über die grafische Oberfläche angesprochen werden muss.

5.2 Gewählte Technologie der Implementierung

Es wird zunächst die bestehende Anwendung von Fooxx betrachtet. Dort werden serverseitig im Bereich der bestehenden Trefferabfrage und -sortierung *ASP*-Scripte eingesetzt, die über *ODBC* mit einer *MySQL*-Datenbank kommunizieren. Als Webserver wird der *Internet Information Server* (*IIS*) unter einem *Windows 2000* System betrieben. Als Technologie für die Implementierung des Algorithmus sowie der Testumgebung ist es sinnvoll, eine Technologie zu wählen, die mit den

Technologien des bestehenden Systems zusammenarbeitet und in dieses möglichst einfach zu integrieren ist. Dies ist insbesondere in Hinsicht auf geplante Portierung des Algorithmus in das bestehende System sinnvoll. Da im Bereich von Suchmaschinen der Faktor Zeit eine wichtige Rolle spielt, ist ferner eine Technologie notwendig, die im Serverbereich eine hohe Leistung erzielt. Vorteilhaft wäre zudem eine Technologie, die ohne hohen Aufwand zu warten und zu erweitern ist.

Es bietet sich zunächst die Verwendung der *ASP*-Technologie selbst an. *ASP* wird hauptsächlich für die Bereitstellung und Verarbeitung von dynamischen Web-Seiten genutzt. Grundsätzlich ist die Implementierung des Algorithmus direkt in *ASP* möglich, bringt aber mehrere entscheidende Nachteile mit sich: Die komplette Testumgebung wäre auf Server-Client Basis zu implementieren, die Nachteile der geringeren Sprachgewalt einer Script-Sprache müssten hingenommen werden und in Bezug auf die Erweiterbarkeit ist die nur teilweise vorhandene Möglichkeit der objekt-orientierten Programmierung zu beachten. Als Vorteil ist dagegen die recht einfache Portierung ins bestehende System nach der fertigen Implementierung zu werten. Ein weiterer Vorteil ist, dass der *IIS*-Server und die *ASP*-Technologie von Microsoft zusammen entwickelt wurden und das Zusammenspiel beider Komponenten hoch entwickelt und ausgereift ist. Somit ist eine hohe Geschwindigkeit bei der Abarbeitung der Scripte zu erwarten. Aufgrund der Erfahrungen des Autors mit der Einschränkung der Scriptsprache *ASP* werden aber weitere Alternativen betrachtet, die eine hohe Kompatibilität zu *ASP* besitzen.

Bei Betrachtung von weiteren kompatiblen Technologien kommt zunächst das *Component Object Model* (*COM*) in Frage. Dieses Modell bietet die Möglichkeit, Klassenbibliotheken in *ASP* zu integrieren. Die Klassenbibliotheken selbst können dabei in jeder zu *COM* kompatiblen Programmiersprache entwickelt werden. Dies sind u.a. *C, C++* und *C#*. Zudem gibt es die Möglichkeit, Teile der bestehenden *ASP*-Scripte in die *ASP.NET* –Technologie zu konvertieren. *ASP.NET* Scripte sind in der Lage, neben *COM* – Objekten auch *.NET*-Objekte zu integrieren. Da es zur Zeit schon Planungen gibt, die bestehende Anwendung zukünftig auch auf eine *.NET*-kompatible Basis zu stellen, bietet es sich an, eine Programmiersprache zu wählen, die COM sowie die *.NET*-Technologie unterstützt, da dadurch die Portierbarkeit in das bestehende System in beiden Fällen gegeben ist.

Die Wahl der Technologie ist auf *.NET* mit der Programmiersprache *C#* gefallen, da in *C#* geschrieben Module selbst *.NET*-Objekte darstellen, in *COM*-Objekte konvertierbar sind und eine hohe Sprachgewalt einer objekt-orientierten Programmiersprache besitzen. Somit ist eine einfache Portierbarkeit in das bestehende System immer gegeben. Weiterhin handelt es sich bei der *.NET*-Technologie um eine Microsoft-Entwicklung, die einen hohen Integrationsgrad mit dem *IIS*-Server unter Windows hat, somit eine hohe Leistung zu erwarten ist und durch die streng objekt-orientierte Struktur einfach zu warten und zu erweitern ist.

5.3 Beschreibung der Anwendungsfälle

Im Folgenden werden die konkreten Anwendungsfälle der Anwendung identifiziert, um eine bessere Aussage über die grundsätzlichen Eigenschaften der Implementierung machen zu können. Es gibt insgesamt vier Anwendungsfälle und einen Akteur, welche wie folgt abgebildet werden können (Abb. 2):

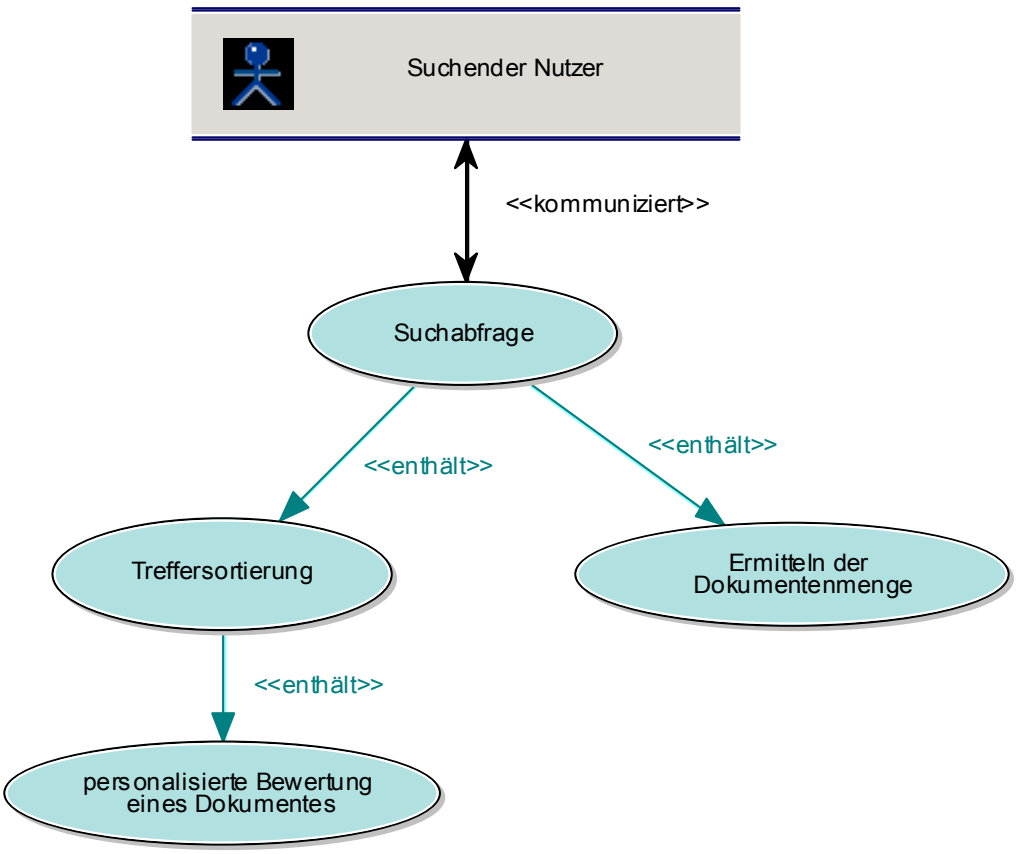


Abbildung 2: Anwendungsfalldiagramm einer Suchabfrage eines Suchenden Nutzers sowie der Sortierung und Personalisierung der relevanten Dokumentenmenge

I. Es gibt einen Anwendungsfall „Suchabfrage“, der den gesamten Ablauf einer Suche und die anschließende Sortierung in einer Dokumentenmenge repräsentiert. Dieser wird durch den Suchenden Nutzer initialisiert, welcher die Suchparameter übergibt und die sortierte Menge an gefundenen relevanten Dokumenten zu seiner Suchabfrage zurück erhält. Innerhalb der Suchabfrage gibt es zwei interne Anwendungsfälle des Systems: „Ermitteln der Dokumentenmenge“ und „Treffersortierung“.

II. Der Anwendungsfall „Ermitteln der Dokumentenmenge“ bezieht sich auf das Finden der relevanten Dokumente der Suchabfrage. Dies kann z.B. durch das Abfragen einer externen Datenquelle mit Hilfe der übergebenen Suchparameter geschehen. Zurückgegeben wird eine Dokumentenmenge mit den relevanten Dokumenten, die zu der Suchabfrage gefunden wurden.

III. Der Anwendungsfall „Treffersortierung“ erhält eine Menge an relevanten Dokumenten und liefert diese sortiert zurück. Dazu wird z.B. der in Kapitel 2 beschriebene Algorithmus genutzt. Der Anwendungsfall Treffersortierung enthält einen weiteren Anwendungsfall „personalisierte Bewertung eines Dokumentes“.

IV. Die „personalisierte Bewertung eines Dokumentes“ liefert nach Übergabe der Ähnlichkeitswerte aller Nutzer im Verhältnis zum Suchenden für ein bestimmtes Dokument den beschreibenden Wert des Dokumentes zurück.

5.4 Ablaufdiagramm

Nachdem die Anwendungsfälle identifiziert wurden, kann aufgrund der gegebenen Informationen ein Ablaufdiagram für eine Testumgebung erstellt werden. Dazu wird der Ablauf eines Testdurchlaufes schematisch dargestellt. (Abb. 3)

Beim Einstiegspunkt der Anwendung wird entschieden, ob die Testumgebung mit oder ohne einer grafischen Oberfläche (der sog. *GUI*) angezeigt werden soll. Anschließend werden die vom Test übergebenden Testparameter eingelesen. Diese Testparameter enthalten Informationen über den suchenden Nutzer, die Suchabfrage selbst sowie Informationen über notwendige Parameter für den

Bewertungsalgorithmus. Sind alle Testparameter vollständig, wird die Dokumentenmenge der relevanten Dokumente ermittelt. Diese kann entweder von außen durch Abfrage einer externen Schnittstelle, z.B. mit Hilfe der von Google bereitgestellten API[10], oder aber von dem Test selbst bereitgestellt werden. Wurden relevante Dokumente gefunden, so wird der Bewertungsalgorithmus mit den entsprechenden Parametern initialisiert. Es werden alle Dokumente durchlaufen: für jedes Dokument werden die erforderlichen Ähnlichkeitswerte im Verhältnis zum suchenden Nutzer aus der entsprechenden Datenquelle abgefragt. Diese Datenquelle kann z.B. die Datenbank der bestehenden Anwendung sein. Anhand dieser Daten wird mit Hilfe des Bewertungsalgorithmus der beschreibende Wert des Dokumentes ermittelt. Wurden alle Dokumente bewertet, wird die Dokumentenmenge entsprechend neu sortiert. Die neu sortierte Dokumentenmenge wird zurückgegeben und für den Fall einer existierenden grafischen Oberfläche zudem entsprechend angezeigt. (Abb. 3)

[10] http://www.google.com/apis/

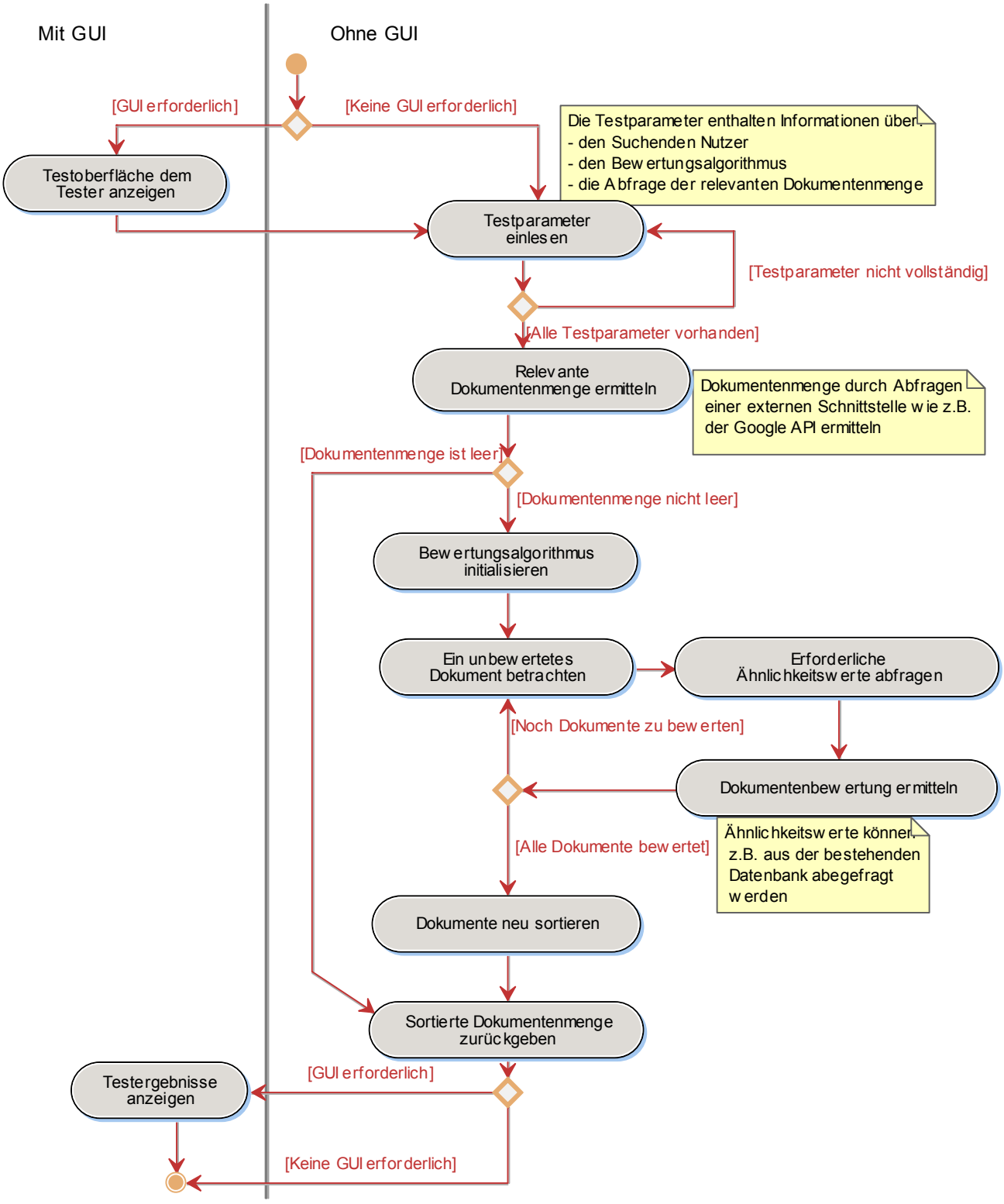


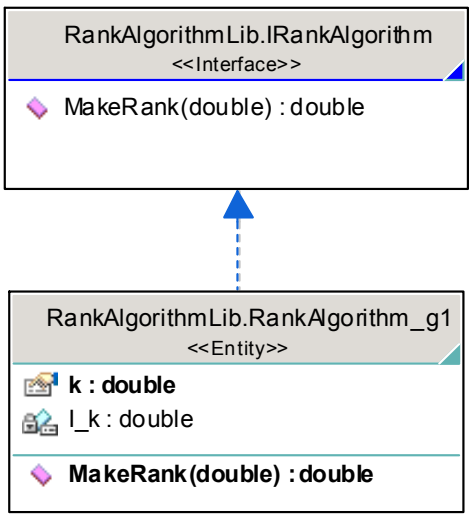
Abbildung 3: Ablaufdiagramm eines Test innerhalb der Testumgebung

5.5 Klassenbeschreibung des Relevanzbewertungs-Algorithmus

Es werden zunächst die benötigten Klassen für die Beschreibung des Personalisierungs-Algorithmus vorgestellt. Diese werden dem Package `RankAlgorithmLib` zugeordnet, welches alle relevanten Klassen für die Algorithmen zum Ermitteln des beschreibenden Wertes beinhalten soll.

Aufgrund von Anforderung 1 wird die Darstellung einer Funktion benötigt, welche die im Kapitel 4.7 vorgestellte Funktion g_1 abbildet. Dazu wird die Klasse `RankAlgorithm_g1` eingeführt. Sie implementiert die Funktion `MakeRank`, welche nach Übergabe der Ähnlichkeitswerte den beschreibenden Wert zurückliefert. Weiterhin besitzt die Klasse die Eigenschaft `k`, welche den im Kapitel 4.7 beschriebenen, gleichnamigen konstanten Faktor k repräsentiert.

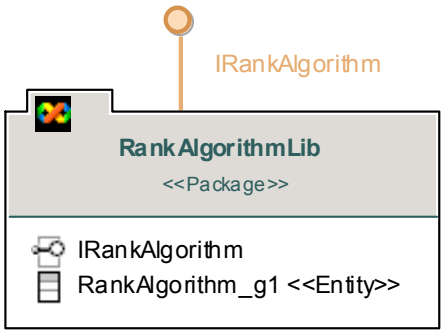
Es muss nach Anforderung 3 eine Schnittstelle definiert sein, über die auf alle Funktionen zugegriffen werden kann. Dies wird durch die Erschaffung der Schnittstelle `IRankAlgorithm` realisiert, welche die Methode `MakeRank` beinhaltet. Somit ist gewährleistet, dass alle Klassen durch Implementation dieser Schnittstelle eine Funktion zur Ermittlung eines beschreibenden Wertes besitzen.
Es ergibt sich folgende schematische Abbildung:



MakeRank(double[] SimilarityValues) : double
Nach Übergabe einer Menge von Ähnlichkeitswerten wird aus diesen der beschreibende Wert nach der Funktion g_1 ermittelt.

k:double
Repräsentiert den im Theorieteil beschriebenen gleichnamigen Faktor. Intern wird die Speicherung in dem Attribut `I_k : double` vorgenommen.

Die mit `I_` beginnenden Attribute stellen jeweils die klasseninterne Speicherung der Eigenschaften dar.

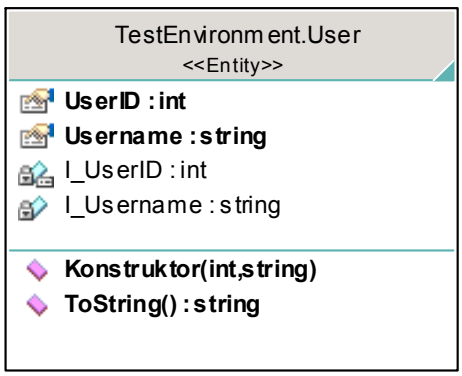


Durch Zusammenfassen dieser Klassen in dem Package `RankAlgorithmLib` kann Anforderung 2 erfüllt werden. Das Package wird als vollwertige *.NET*-Bibiliothek angelegt und kann zudem noch zu einem *COM*-Objekt kompiliert werden. Dadurch wird es in die bestehende Anwendung portierbar sein.

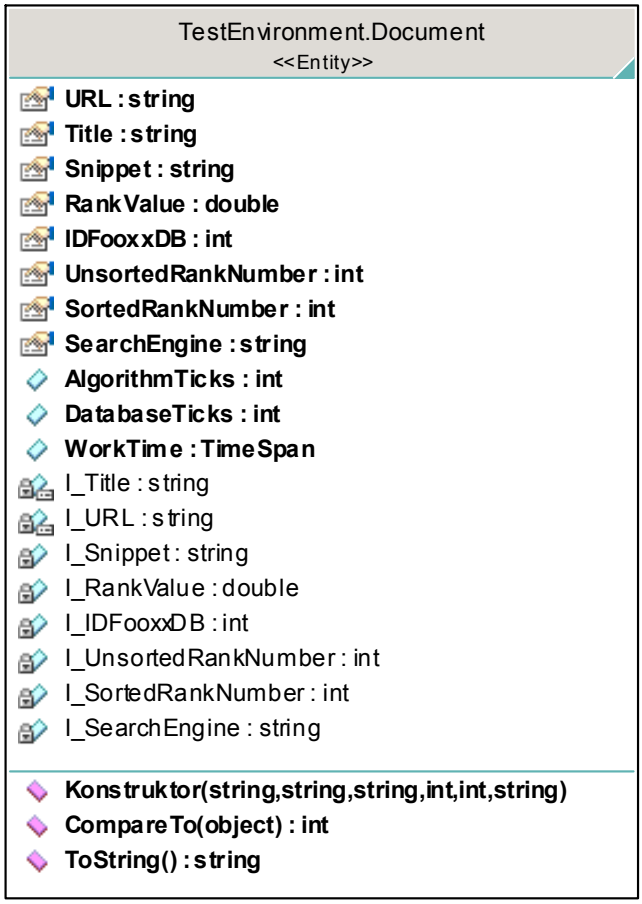
5.6 Klassenbeschreibung für die Testumgebung

Es werden die Klassen für die Testumgebung identifiziert, welche dem Package `TestEnvironment` zugeordnet werden.

Zunächst werden in der Testumgebung Klassen für die Repräsentation der Begriffe des Dokumentes und des Nutzers benötigt, um eine Abbildung der Theorie auf die Praxis umsetzen zu können. Sie repräsentieren prinzipiell nur Datentypen mit grundlegenden Funktionen. Die beiden Klassen `Document` und `User` stellen jeweils ein Dokument (in der Regel eine Webseite) und einen Nutzer dar. Damit verschiedene Dokumente anhand ihres beschreibenden Wertes miteinander verglichen werden können, implementiert `Document` die Schnittstelle `IComparable`. Um Anforderung 6 zu erfüllen, können in einem Dokument zudem verschiedene Zeitmessungen gespeichert werden. Im Folgenden die schematische Abbildung von `User` und `Document`:



UserID : int
Eindeutiger Integerwert zur Identifizierung des Nutzers.
Username : string
Name des Nutzers.
Konstruktor User (int UserID, string Username)
Mit Angabe der User-ID und dem Namen des Nutzers wird ein neuer Nutzer angelegt.
ToString () : string
Gibt eine Zeichenkette mit Informationen über den Nutzer zurück.



URL : string

Die URL dient zur eindeutigen Identifizierung des Dokumentes.

Title : string

Speichert den Titel des Dokumentes.

Snippet : string

Gibt eine kurze Beschreibung des Dokumentes.

RankValue : double

Speichert den vom Relevanzbewertungs-Algorithmus g_1 ermittelten beschreibenden Wert des Dokumentes.

IDFoxxDB : int

Interne ID des Dokumentes in der Fooxx-Datenbank.

UnsortedRankNumber : int

Rang in der unsortierten Dokumentenmenge vor der Sortierung.

SortedRankNumber : int

Rang in der sortierten Dokumentenmenge.

SearchEngine : string

Speichert, mit welcher Suchmaschine das Dokument gefunden wurde.

AlgorithmTicks : int

Anzahl der Prozessor-Ticks, die zur Ermittlung des beschreibenden Wertes für dieses Dokument verstrichen sind.

WorkTime : TimeSpan

Zeitspanne, die insgesamt benötigt wurde, um alle Ähnlichkeitswerte abzufragen und den beschreibenden Wert für dieses Dokument zu ermitteln.

DatabaseTicks : int

Anzahl der Prozessor-Ticks, die zur Abfrage der Ähnlichkeitswerte aus der Datenbank benötigt wurden.

Konstruktor Document (string URL, string Title, string Snippet, int IDFooxxDB, int UnsortedRankNumber, string SearchEngine)

Nach Übergabe der URL, des Titels, dem Snippet, der internen Datenbank-ID, dem Rang in der unsortierten Dokumentenmenge und der abgefragten Suchmaschine des Dokumentes wird ein neues Dokument angelegt.

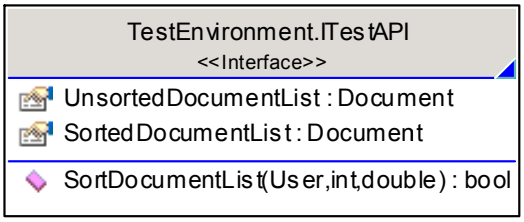
CompareTo (object Obj) : int

Überladene Methode von IComparable – sie implementiert die Möglichkeit, Dokumente mit dem vom Relevanzbewertungs-Algorithmus g_1 ermittelten beschreibenden Wert (dem `RankValue`) zu vergleichen und zu sortieren.

ToString () : string

Gibt eine Zeichenkette mit Informationen über das Dokument zurück.

Nach Anforderung 8 muss die Testumgebung eine Schnittstelle bereitstellen, über die auf die Dokumentenmenge zugegriffen werden kann. Sie dient dazu, die bereits in dem Anwendungsfall „Testdurchführung“ beschriebene Übergabe einer unsortierten Dokumentenmenge und Zurückgabe einer sortierten Dokumentenmenge abzuwickeln. Diese Schnittstelle muss somit die Möglichkeit bieten, eine Dokumentenmenge zu empfangen, diese zu sortieren und die sortierte Dokumentenmenge zugänglich zu machen. Die Folgende schematisch dargestellte Schnittstelle `ITestAPI` repräsentiert diesen Sachverhalt:



UnsortedDocumentList[] : Document

In diesem Feld wird die abgefragte und unsortierte Dokumentenmenge vorgehalten.

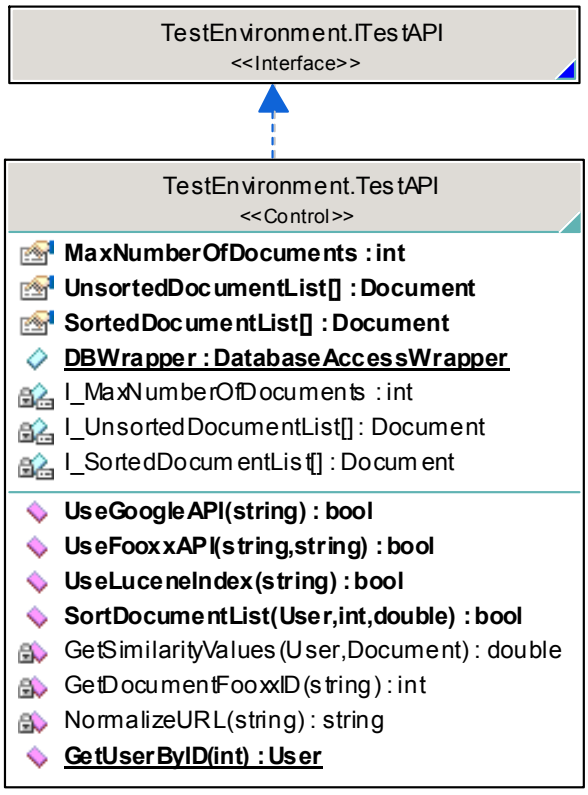
SortedDocumentList[] : Document

In diesem Feld wird nach der Sortierung die sortierte Dokumentenmenge vorgehalten.

sortDocumentList (User SearchUser, int AlgorithmNumber, params double [ ] AlgorithmParameter) : bool

Erwartet als Parameter den Suchenden Nutzer vom Typ `User`, die Nummer des zu verwendenden Algorithmus (z.B. 1 für `RankAlgorithm_g1`) sowie eine Liste mit evtl. notwendigen Parametern für diesen Algorithmus. Liefert als Rückgabewert `true` zurück, falls das Sortieren erfolgreich war, ansonsten `false`.

Für die Testumgebung ist weiterhin eine Steuerklasse notwendig, um den zeitlichen Ablauf und die Koordination der einzelnen Komponenten zu kontrollieren. Es werden in dieser Klasse feststehende Konfigurations-Parameter für die gesamte Testumgebung festgelegt. Diese Klasse implementiert die beschriebene Schnittstelle `ITestAPI`. Die Steuerklasse hat den folgenden schematischen Aufbau:



MaxNumberOfDocuments : int

Maximale Anzahl an Dokumenten der Dokumentenmenge.

UseGoogleAPI (string Searchstring) : bool

Nach Übergabe eines Suchbegriffes wird die Google-API benutzt, um die unsortierte Dokumentliste mit Dokumenten zu füllen. Bei erfolgreicher Durchführung wird `true` zurückgeliefert.

UseFooxxAPI (string Searchstring) : bool
Nach Übergabe eines Suchbegriffes wird die Fooxx-API benutzt, um die unsortierte Dokumentenliste mit Dokumenten zu füllen. Bei erfolgreicher Durchführung wird `true` zurückgeliefert.

UseLuceneIndex (string Searchstring) : bool
Nach Übergabe eines Suchbegriffes wird der eigene Index benutzt, um die unsortierte Dokumentenliste mit Dokumenten zu füllen. Bei erfolgreicher Durchführung wird `true` zurückgeliefert.

GetSimiliarityValues (User Searchuser, Document Document) : double[]
Liefert für die übergebenen Dokumente alle Ähnlichkeitswerte der anderen Nutzer im Verhältnis zum übergebenen Nutzer zurück.

GetDocumentFooxxID (string URL) : int
Liefert für die übergebene URL die (falls vorhanden) ID des Dokumentes in der Fooxx-Datenbank zurück.

NormalizeURL (string URL) : string
Hilfsfunktion, welche die übergebene URL auf eine einheitliche Form bringt und diese zurückgibt.

GetUserByID (int UserID) : User
Hilfsfunktion, die für eine übergebene ID den entsprechenden Nutzer zurückgibt.

Durch die Implementierung der Methode `UseGoogleAPI` kann über eine von Google bereit gestellte Schnittstelle die unsortierte Dokumentenmenge von der Testumgebung selbst abgefragt werden. Ebenso werden mit den Methoden `UseFooxxAPI` und `UseLuceneIndex` jeweils die bereitgestellten Schnittstellen von Fooxx sowie vom eigenen Index von Fooxx abgefragt. Dies erfüllt Anforderung 5, dass die Testumgebung selbst eine unsortierte Dokumentenmenge durch das Abfragen von Suchmaschinen bereitstellen soll.

In der Steuerklasse kann bei Bedarf die grafische Oberfläche `GUI` gestartet werden, die nach Anforderung 7 für den Testverlauf notwendig ist. Die grafische Oberfläche ist fest mit der Steuerklasse verbunden und bietet eine Schnittstelle zur Kommunikation zwischen Tester und Testdurchführung. Des Weiteren muss die grafische Oberfläche noch die Möglichkeit bieten, die Nutzer auszuwählen und die Dokumente einzeln zu betrachten. Dies wird in den Klassen `ChooseUserDialog` sowie `ViewDocument` realisiert, wobei die Klasse `ChooseUserDialog` über die Methode `GetUserByID` direkt mit der Testumgebung kommuniziert. Die grafische Oberfläche besitzt den folgenden Aufbau (Abb. 4):

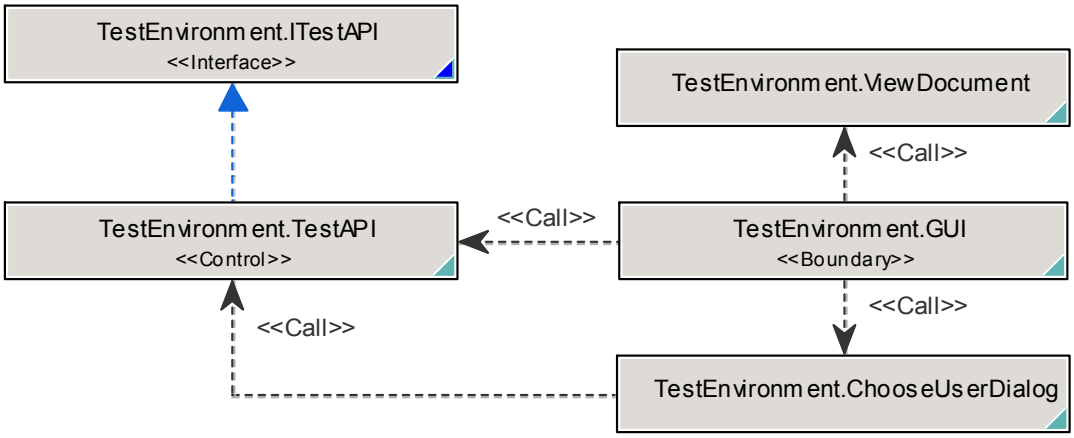
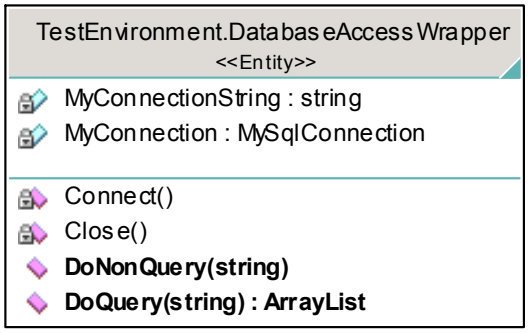


Abbildung 4: Klassenmodell der grafischen Oberfläche

Abbildung 4 verdeutlicht die Struktur der grafischen Oberfläche, die exakte Abbildung der wirklich notwendigen Methoden, Attribute und Eigenschaften werden hier nicht näher erläutert. Die Anzahl aller Komponenten bei grafischen Oberflächen ist recht groß und kann erst nach Entwurf des Oberflächendesigns konkret formuliert werden. Wichtig ist hierbei, dass die grafische Oberfläche die unsortierte sowie sortierte Dokumentenmenge anzeigen kann. Zudem kann eine Dokumentenmenge geladen und die Personalisierung der Dokumentenmenge vom Tester über die Oberfläche initialisiert werden.

Damit die Testumgebung Zugang zur Datenbank erhält, um damit erforderliche Daten für den Relevanzbewertungs-Algorithmus bereitzustellen und Anforderung 4 gerecht zu werden, wird die Klasse `DatabaseAccessWrapper` eingeführt. Sie bietet der Testumgebung die Möglichkeit, auf die Datenbank zuzugreifen. `DatabaseAccessWrapper` ist eine Klasse, die nur die grundlegenden Funktionen zur Verbindung mit der Datenbank bereitstellt – es können Queries an die Datenbank geschickt und die eventuelle Antwort gelesen werden. Dabei bezeichnet `DoNonQuery` einen Query ohne Ergebnis und `DoQuery` einen Query mit Ergebnis. Die Herstellung und das Beenden einer Verbindung mit `Connect` bzw. `Close` wird innerhalb der Methoden `DoNonQuery` und `DoQuery` aufgerufen und sind deswegen als nicht öffentlich deklariert. Diese Klasse stellt die Verbindung zwischen der bestehenden *MySQL*-Datenbank und der Testumgebung bereit.



MyConnectionString : string

Zeichenkette, die die notwendigen Informationen für die Verbindung enthält.

MyConnection : MySqlConnection

Objekt, welches die Datenbankverbindung repräsentiert.

Connect ()

Stellt eine Verbindung mit der Datenbank her.

Close ()

Beendet die Verbindung mit der Datenbank.

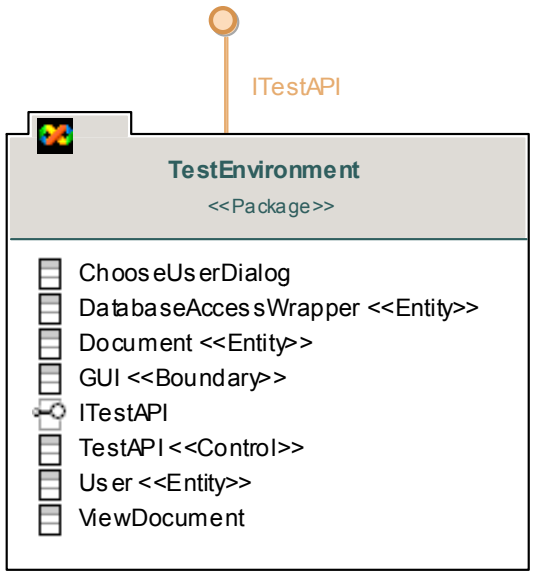
DoNonQuery (string NonQuery)

Führt das übergebende Query aus. Das Query darf keine Ergebnisse liefern.

DoQuery (string Query) : ArrayList

Führt das übergebende Query aus. Die Ergebnisse werden mit der Klasse `System.Collections.ArrayList` übergeben, welches ein dynamisch wachsendes Feld repräsentiert und Zugriff auf den Inhalt ähnlich einem Array fester Größe gewährt.

Nach dem alle Anforderungen an die Testumgebung erfüllt sind, kann das Package `TestEnvironment` erstellt werden. Nach der Entwicklung des Packages für die Testumgebung wird nun das Gesamtklassenmodell näher betrachtet.



5.7 Gesamtklassenmodell

Das Gesamtklassenmodell verdeutlicht die Beziehungen aller Klassen untereinander. Die Grafische Oberfläche `GUI` greift dabei auf die Steuerklasse `TestAPI` sowie auf die Dokumente `Document` und den Nutzer `User` zu und interagiert mit diesen. Die Steuerklasse der Testumgebung `TestAPI` interagiert mit den entsprechenden Schnittstellen von Google, Fooxx sowie des eigenen Index. Zudem stellt die Steuerklasse die Verbindung mit der Datenbank über den `DatabaseAccessWrapper` her. Nach der Abfrage einer unsortierten Dokumentenmenge und den entsprechenden Daten aus der Datenbank wird mit dem Relevanzbewertungs-Algorithmus die unsortierte Dokumentenmenge bewertet, sortiert und entsprechend zur Anzeige gebracht. (Abb. 5)

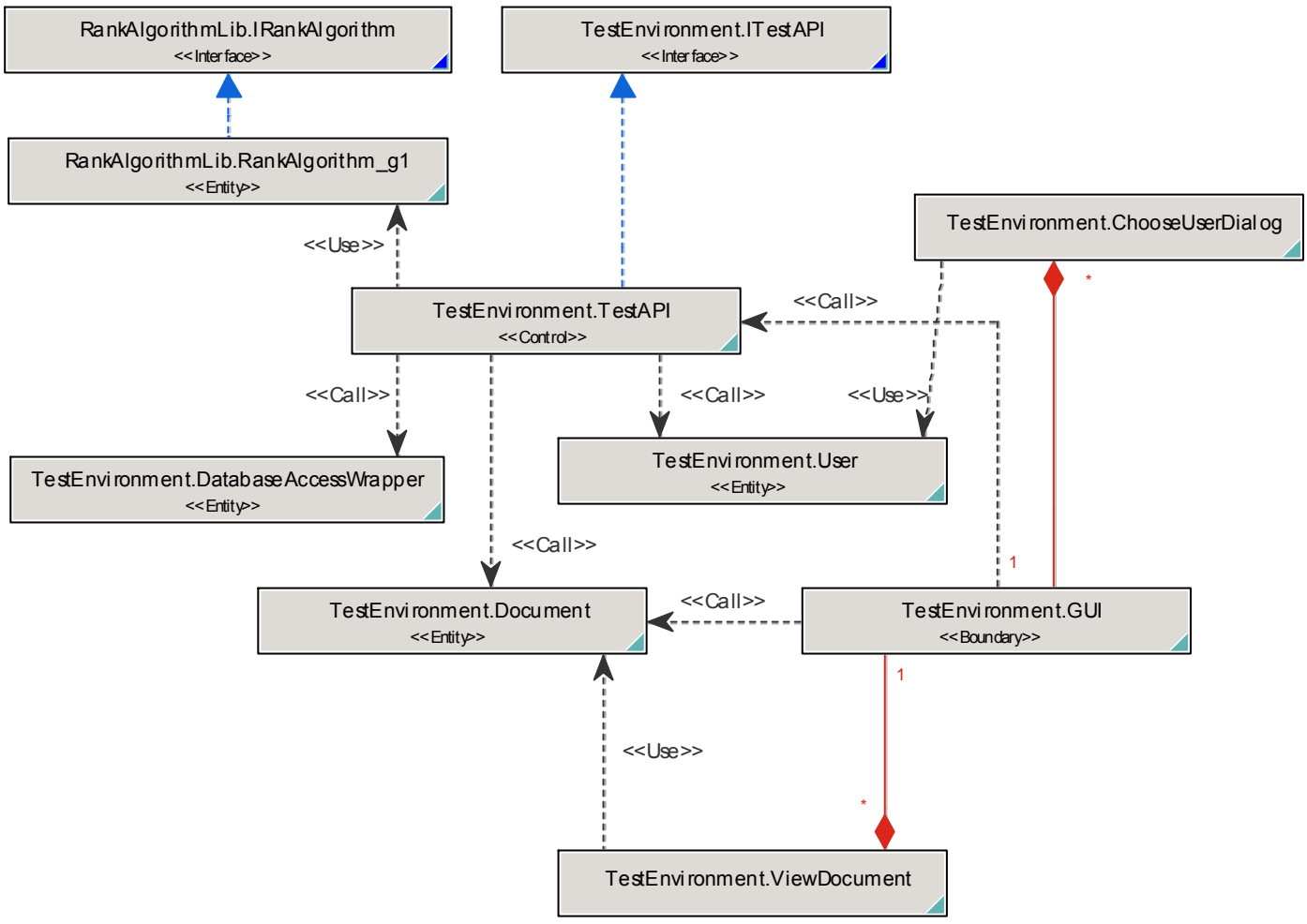


Abbildung 5: Gesamtklassenmodell der Implementierung

5.8 Oberflächendesign

Im Folgenden werden die Oberflächen der Anwendung beschrieben.

5.8.1 Hauptfenster der Testumgebung

Das Hauptfenster der Anwendung, über welches diese gesteuert werden kann, wird durch die Klasse `GUI` repräsentiert. (Abb. 6)

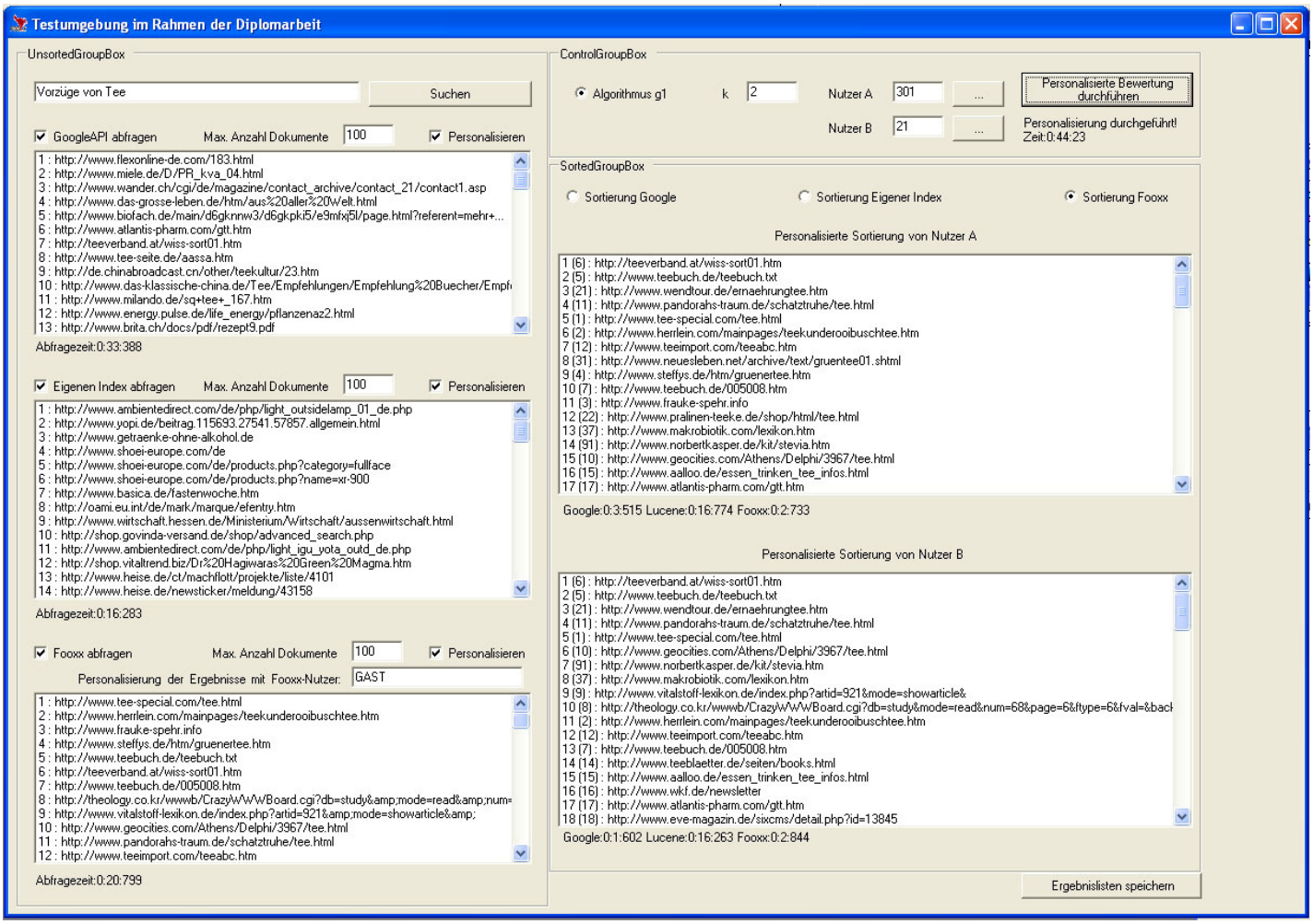


Abbildung 6: Screenshot des Hauptfensters der Testumgebung

Die Oberfläche der Testumgebung (Abb. 6) besteht aus drei Bereichen. Dabei befindet sich Bereich 1 auf der linken Seite. Er besteht hauptsächlich aus drei großen sowie einigen kleineren Textfeldern. Bereich 2 ist in der rechten oberen Ecke zu sehen und besteht aus mehren kleineren Einstellungsfeldern sowie einem großen Knopf zum Starten der Personalisierung. Bereich 3 entspricht dem restlichen rechten Abschnitt und besteht hauptsächlich aus zwei großen Textfeldern. Die einzelnen Bereiche werden nun näher erläutert.

Bereich 1: Der linke Teil der Testumgebung dient zur Abfrage der unsortierten Dokumentenmenge. Dies kann einmal über die von Google bereitgestellte Schnittstelle erfolgen. Weiterhin kann der eigene Index abgefragt werden. Die dritte Möglichkeit besteht in der Abfrage der Dokumentenmenge von der bereitgestellten Schnittstelle von Fooxx. Der Suchbegriff, mit denen diese drei Schnittstellen abgefragt werden, wird im oberen Eingabefeld des Bereichs eingegeben. Durch Klicken auf den Knopf „Suchen“ rechts neben dem Eingabefeld für den Suchbegriff werden die Abfragen nacheinander durchgeführt. Es ist möglich, die einzelnen Abfragen für die Suchmaschinen gezielt an/abzustellen sowie die maximale Anzahl an abzufragenden Dokumenten einzugeben. Zusätzlich kann für die Suchmaschine Fooxx noch ein Nutzer angegeben werden, für den eine eventuelle personalisierte Sortierung der Ergebnisse erfolgen soll. Nach der Durchführung einer Abfrage werden die Dokumente und die Abfragezeit in den entsprechenden Feldern angezeigt.

Bereich 2: im rechten oberen Bereich der Testumgebung können Einstellungen zu dem Relevanzbewertungs-Algorithmus gemacht werden. Es werden dazu alle implementierten Algorithmen angezeigt. Im Rahmen dieser Arbeit steht ausschließlich der in Kapitel 4.7 beschriebene Bewertungsalgorithmus g_1 zur Verfügung. Zusätzlich können weitere Parameter für diesen Algorithmus festgelegt werden. Dies ist für den Bewertungsalgorithmus g_1 der Parameter *k*. Weiterhin können zwei suchende Nutzer A und B angegeben werden, für die die Dokumentenmenge personalisiert werden sollen. Der Nutzer kann entweder direkt durch Eingeben der ID des Nutzers oder durch Klicken auf den Knopf „…“ einen Nutzer auswählen. Wird auf „…“ geklickt, so kommt man in das unter 5.8.2 beschriebene Auswahlmenü. Der Knopf „Personalisierte Bewertung durchführen“ startet die Durchführung der Bewertung und Sortierung der Dokumente für die angegebenen Nutzer. Die im Bereich 1 von den Suchmaschinen abgefragten Dokumente werden personalisiert und die Ergebnisse in Bereich 3 angezeigt. Die gesamte Dauer der Personalisierung wird unter dem Knopf „Personalisierte Bewertung durchführen“ angezeigt.

Bereich 3: In diesem Bereich werden die personalisierten Dokumentenmengen der abgefragten Suchmaschinen angezeigt. Es existierten jeweils zwei Felder zur

Anzeige der personalisierten Dokumentenmengen – das obere Feld für die personalisierte Dokumentenmenge des Nutzers A und das untere für den Nutzer B. Dabei wird durch Klicken der sog. „Radiobuttons“ im oberen Teil die entsprechende personalisierte Dokumentenmenge in den Feldern für die Nutzer A und B angezeigt. Die Dauer der Personalisierung der einzelnen abgefragten Suchmaschinen erscheint unter den Feldern.

Für die Bereiche 1 und 3 gilt: Die Dokumente werden jeweils durch Ihre URLs repräsentiert. Durch Doppelklick auf eine der angezeigten URLs eines Dokumentes wird eine neues Fenster geöffnet, welches weitere Informationen über das Dokument enthält und unter 5.8.3 beschrieben wird.

Rechts unten in der Testumgebung befindet sich noch ein weiterer Knopf mit der Beschriftung: „Ergebnislisten speichern“. Klickt man auf diesen, kann man nach Angabe einer Datei die angezeigten Ergebnislisten im Text-Format speichern. Beispiele für das Format dieser Dateien finden sich im Anhang unter B18, B19, B20, B21 und B22.

5.8.2 Fenster nach Klicken auf „…“

Dieses Fenster wird durch die Klasse `ChooseUserDialog` repräsentiert.

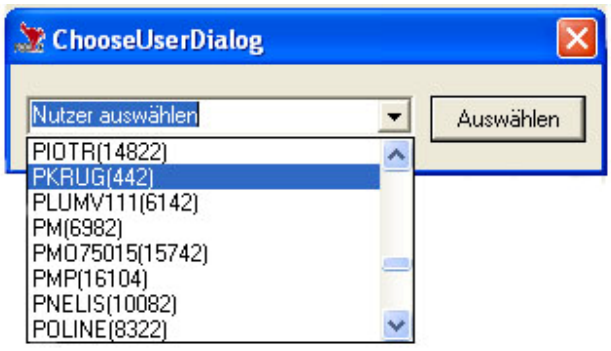


Abbildung 7: Screenshot des Dialogfeldes nach Klicken auf „…“ im Hauptfenster

In dem angezeigten „Pull-down-Menü“ (Abb. 7) kann man einen der alphabetisch geordneten Nutzer von Fooxx auswählen. Nach dem Drücken des Knopfes „Auswählen“ wird die ID des ausgewählten Nutzers in den entsprechenden Bereich des Hauptfensters übertragen. Die Anwendung wird solange blockiert, bis ein Nutzer ausgewählt wurde.

5.8.3 Fenster nach Doppelklick eines Dokumentes

Dieses Fenster wird durch die Klasse `ViewDocument` dargestellt.

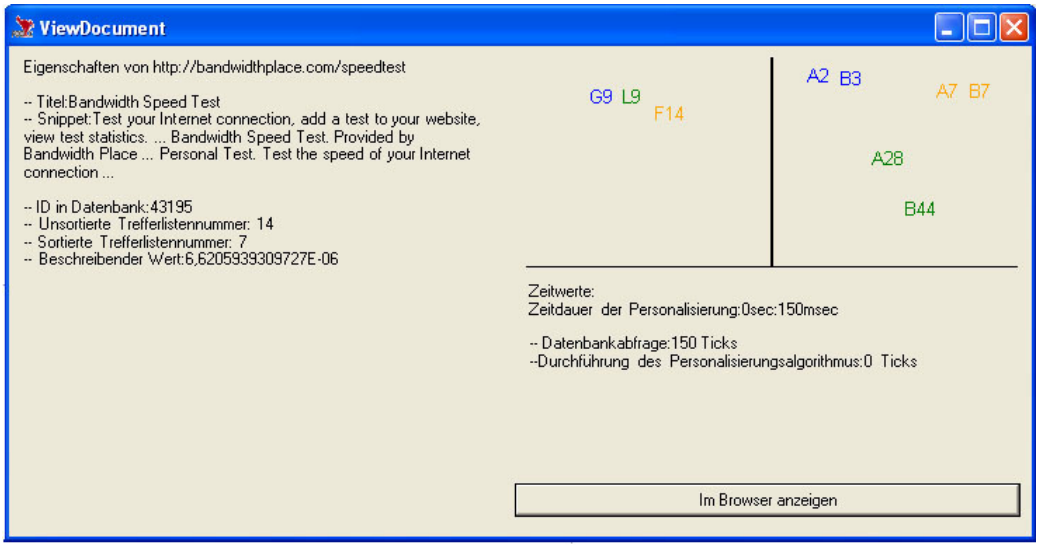


Abbildung 8: Screenshot des Fensters nach Doppelklicken auf ein Dokument im Hauptfenster

In diesem Fenster werden alle wichtigen Informationen über das Dokument, welches sich durch seine URL eindeutig identifiziert, angezeigt. Dazu gehören: der Titel, ggf. ein Snippet, die ID des Dokumentes in der Fooxx-Datenbank (falls vorhanden), der Rang in der unsortierten Dokumentenmenge, den Rang in der personalisierten Dokumentenmenge (falls vorhanden) und nach der Personalisierung noch Zeitwerte zur Dauer der Personalisierung. Außerdem ist eine grafische Übersicht über den Verlauf des Dokumentes in den einzelnen Suchmaschinen vor und nach der Personalisierung rechts im Fenster abgebildet. Es kann nun die Implementierung der Testumgebung vorgenommen werden.

5.9 Fertige Implementierung

Nach der Umsetzung des Entwurfs entstand die fertige Implementierung der Testumgebung. Da sie Daten aus der Fooxx-Datenbank und aus dem Internet abfragt, benötigt sie zum fehlerfreien Betrieb Zugriff auf beide Quellen. Weiterhin ist für die Ausführung das *.NET*-Framework notwendig, welches bei Microsoft für Windows-Systeme kostenlos herunter geladen werden kann. Ab Windows XP Service Pack 1 ist es bereits im Betriebssystem integriert.

Der komplette Quellcode befindet sich im Anhang unter A4. Er umfasst über 2000 Zeilen Code. Der Ort der ausführbaren Dateien ist im Anhang unter A5 beschrieben. Es folgen nun Erläuterungen zum Quellcode der Implementierung.

5.10 Erläuterungen zum Quellcode

Nachdem die Implementierung der Testumgebung stattgefunden hat, werden im Folgenden einzelne Teile des Quellcodes besprochen, die im Bezug zu dieser Arbeit interessant sind.

5.10.1 Implementierung des Relevanzbewertungs-Algorithmus

Zunächst wird die Implementierung des Relevanzbewertungs-Algorithmus betrachtet. Der Quellcode hat folgenden Aufbau: (Quellcode 1)

```
public class RankAlgorithm_g1 : IRankAlgorithm
{
	/// <summary>
	/// Variabler Faktor zur Einflussnahme auf die Bedeutung von
	/// Ähnlichkeitswerten.
	/// </summary>
	public double k
	{
		get
		{
			return I_k;
		}

		set
		{
			I_k = value;
		}
	}
	private double I_k;
	/// <summary>
	/// Berechnet nach Übergabe der Ähnlichkeitswerte den beschreibenden Wert
	/// eines Dokumentes.
	/// </summary>
	/// <param name="SimilarityValues">Feld der Ähnlichkeitswerte</param>
	/// <returns>Beschreibender Wert, der sich aus den Ähnlichkeitswerten
	/// ergibt</returns>
	public double MakeRank (double[] SimilarityValues)
	{
		double rank=0;
		foreach (double s_value in SimilarityValues)
		{
			rank=rank+Math.Pow(s_value,k);
		}
		rank=1-1/(rank+1);
		return rank;
	}
}
```

Quellcode 1: Quellcode des Relevanzbewertungs-Algorithmus RankAlgorithm_g1

Der Relevanzbewertungs-Algorithmus g_1, welcher in 4.7 dargestellt wurde, wird in der Klasse `RankAlgorithm_g1` abgebildet. Der Quellcode der Klasse ist in Quellcode 1 abgebildet. Die Methode `MakeRank` stellt die eigentliche

Implementierung des Algorithmus dar und erwartet somit als Übergabeparameter ein Feld von Ähnlichkeitswerten des Datentyps `double`. Nach Aufruf der Funktion wird zunächst der beschreibende Wert `rank` mit dem Wert 0 initialisiert. Es werden alle Ähnlichkeitswerte einzeln durchlaufen und aufsummiert, wobei sie vorher mit dem Parameter `k` potenziert werden. `k` muss vor Aufruf der Funktion entsprechend initialisiert worden sein. Nach dem Aufsummieren wird der Kehrwert der Summe plus 1 gebildet und das Ergebnis von 1 abgezogen. Das Endergebnis ist der beschreibende Wert, welcher zurückgeliefert wird.

Der zweite wichtige Punkt für die Umsetzung des Relevanzbewertungs-Algorithmus ist das Abfragen der Ähnlichkeitswerte der Nutzer, welcher elementar für die Berechnung des beschreibenden Wertes eines Dokumentes ist. Dies wird in der Funktion `GetSimiliarityValues` in der Klase `TestAPI` umgesetzt und hat den folgenden Aufbau: (Quellcode 2)

```
private double[] GetSimilarityValues (User searchUser, Document document)
{
      document.URL = NormalizeURL(document.URL);

      string query = "SELECT similarity FROM" +
                  " (SELECT DISTINCT user_vid FROM protocol p, objects o " +
                  "  WHERE p.object_id = o.id AND o.url = '" + document.URL + "')" +
                  "a, similarity b WHERE a.user_vid = b.a_user_id " +
                  "AND b.b_user_id = " + searchUser.UserID;

      ArrayList dbResults = DBWrapper.DoQuery(query);

      double[] s = new double[dbResults.Count];
      int count=0;
      foreach (object[] o in dbResults)
      {
            s[count++] = Double.Parse(o[0].ToString());
      }
      return s;
}
```

Quellcode 2: Methode GetSimilarityValues zur Abfrage der Ähnlichkeitswerte aus der Datenbank

```
SELECT similarity FROM
      (SELECT DISTINCT user_vid FROM protocol p, objects o
       WHERE p.object_id = o.id AND o.url = 'document.URL')
a, similarity b WHERE a.user_vid = b.a_user_id
AND b.b_user_id = " + searchUser.UserID;
```

Quellcode 3: Die Datenbankabfrage aus Quellcode 2 ohne störende Begrenzungszeichen

Die Funktion `GetSimilartiyValues` erwartet als Übergabeparameter einen Nutzer vom Typ `User` sowie ein Dokument von Typ `Document`. Die Funktion

ermittelt für die übergebenen Dokumente alle Ähnlichkeitswerte der Nutzer, die dieses Dokument benutzt haben, wobei die Ähnlichkeitswerte im Verhältnis zum übergebenen Nutzer berechnet werden. Die Datenbankabfrage ist eine einfach verschachtelte *SQL*-Abfrage vom Typ *SELECT* und wird in Quellcode 3 lesbarer dargestellt. Im inneren *SELECT* der Abfrage werden die IDs der Nutzer ermittelt, die das übergebene Dokument genutzt haben. Im äußeren *SELECT* werden dann die Ähnlichkeitswerte der Nutzer dieses Dokumentes im Verhältnis zum Suchenden Nutzer ermittelt. Dabei bezieht sich die SQL-Abfrage auf die in Abbildung 9 dargestellte Tabellenstruktur der Fooxx-Datenbank.

Abbildung 9: Struktur der relevanten Tabellen `protocol`, `objects` und `similarity` der Fooxx-Datenbank

5.10.2 Abfragen der einzelnen Suchmaschinen

In der Klasse `TestAPI` werden mit den Methoden `UseGoogleAPI`, `UseLuceneIndex` und `UseFooxxAPI` Möglichkeiten eingebunden, eine unsortierte Dokumentenmenge für die Testumgebung bereitzustellen. Der Quellcode dazu findet sich im Anhang unter A1, A2 und A3. Es wird nun die Vorgehensweise dieser drei Methoden im Einzelnen detaillierter erläutert.

`UseGoogleApi` (Anhang A1)

Diese Methode greift auf den von Google bereitgestellten Web-Service zu. Die Methode stellt eine Verbindung mit dem Webservice über die von Google angebotene API her. Diese API verbindet sich mit dem Google-Server über eine *SOAP/WSDL* Verbindung. Pro Anfrage können dabei 10 Treffer für einen Suchbegriff angefordert werden. Die Methode `UseGoogleApi` fordert hintereinander die in der

Testumgebung angegebene Anzahl an Treffern von Google an und gibt diese dann an die Testumgebung weiter.

`UseLuceneIndex` (Anhang A2)
Die Methode und die damit verbundene Abfrage eines eigenen Index wurden speziell im Rahmen dieser Arbeit entwickelt. Zunächst wurden die ca. 160.000 bekannten, in der Datenbank von Fooxx vorhandenen Dokumente mit Hilfe des vom Autor entwickelten Tools `ParallelHTTPClient` herunter geladen und als Text-Dateien lokal gespeichert. Dann wurde die *Open-Source*-Volltextsuchmaschine Lucene[11] benutzt, um aus diesen herunter geladenen Dokumenten einen „eigenen" Index für Fooxx zu erstellen. Lucene ist eine Volltextsuchmaschine, welche zunächst in Java entwickelt wurde. Es gibt allerdings bereits Portierungen für die *.NET*-Technologie. Diese Portierung wurde verwendet, um einen eigenen Index zu erstellen. Dieser Index wird dann beim Aufruf der Funktion mit Lucene durchsucht und die relevanten Seiten ermittelt. Diese werden daraufhin an die Testumgebung weitergegeben.

Da die indexierten 160.000 Seiten nur einen sehr kleinen Teil an existierenden Seiten im Internet widerspiegeln und somit nicht unbedingt alle Informationsbereiche für bestimmte Suchanfragen abdecken, liefert die Abfrage des eigenen Index nur bedingt sinnvolle Ergebnisse. Auch die Volltextsuche selbst ist nicht die optimale Vorgehensweise für das Auffinden von relevanten Webseiten, da sie z.B. nicht die Verlinkungsstrukur der Seiten untereinander berücksichtigt.

`UseFooxxIndex` (Anhang A3)
Diese Methode dient dazu, die im Rahmen dieser Arbeit entwickelte Fooxx-API abzufragen. Die Fooxx-Api stellt dabei eine URL dar, die über das *http*-Protokoll abgefragt wird und eine *XML*-Seite mit den entsprechenden Treffern für die übermittelte Suchanfrage zurückliefert.

Nachdem Teile des Quellcodes erläutert wurden, wird nun mit der Testumgebung der entwickelte Relevanzbewertungs-Algorithmus getestet.

[11] http://lucene.apache.org/

6 Durchgeführte Tests und Auswertung

In diesem Teil werden Tests zu Bewertung des im Rahmen dieser Arbeit entwickelten Relevanzbewertungs-Algorithmus mit Hilfe der Testumgebung durchgeführt.

6.1 Testreihe I

Im Rahmen einer öffentlichen Veranstaltung wird von 15 Testpersonen ein vorgelegtes Aufgabenblatt bearbeitet. Dieses Aufgabenblatt verlangt von den Testern, bei einer gegebenen Suchmaschine eine vorgegebene Anfrage einzugeben und 5 Seiten aus der erhaltenen Ergebnisliste als relevant zu bewerten. Dazu sollte der Tester einzelne Seiten der Ergebnisliste betrachten und jene Seiten benennen, die er als relevant für die Suchanfrage einschätzt. Im Aufgabenblatt werden zwei Abfragen mit jeweils zwei Suchmaschinen vorgegeben:

- Die Abfrage „Vorzüge von Tee“ in der Suchmaschine Fooxx
- Die Abfrage „Abholzung Regenwald“ in der Suchmaschine Google

Das Aufgabenblatt, das den Testern vorgelegt wurde, findet sich im Anhang B1.

Die Tester sind während des Tests mit einem bestehenden Nutzer bei der Fooxx-Toolbar angemeldet, welche die Interaktionen des Testers und somit die aufgerufenen Seiten mitprotokolliert. Es stehen vier Nutzer von Fooxx zur Verfügung, die bereits ein Nutzerprofil besitzen. Die Tester werden dabei möglichst gleich verteilt mit diesen bestehenden Fooxx-Nutzern angemeldet.
Die vorgegebenen Nutzer sind:

- DASDINGSA (Fooxx-ID: 2623)
- ROADRUNNERLENNY (Fooxx-ID: 9242)
- VAH (Fooxx-ID: 21)
- JKESSLER (Fooxx-ID: 16137)

Dabei sind die Nutzer DASDINGSDA, VAH und ROADRUNNERLENNY häufige Nutzer, JKESSLER ein seltener Nutzer von Fooxx.

Zwar wäre es möglich gewesen, für die einzelnen Tester jeweils einen eigenen neuen Nutzer anzulegen, allerdings hätte der Test dann keine sinnvolle Ergebnisse gebracht, da für die Anwendung des Algorithmus ein Nutzer mit einem bestehenden Profil Voraussetzung ist (d.h. dem Nutzer müssen eine Anzahl an besuchten

Webseiten zugeordnet sein). Dies wird später an den Testergebnissen des Fooxx-Nutzers JKESSLER deutlich, welcher so gut wie keine Interaktionen an Fooxx übermittelt hat und somit kein vergleichbares Profil besitzt.

Berechnet man die Ähnlichkeit der Nutzer untereinander, betrachtet also die übereinstimmenden Dokumente der einzelnen Nutzer, so erhält man folgende (normalisierte) Werte für die Ähnlichkeit:

	Ähnlichkeiten der Nutzer			
	DASDINGSDA	ROADRUNNERLENNY	VAH	JKESSLER
DASDINGSDA	1	0,00106213	0,00546992	0
ROADRUNNERLENNY	0,00106213	1	0,00233645	0
VAH	0,00546992	0,00233645	1	0
JKESSLER	0	0	0	1

Tabelle 2: Ähnlichkeitswerte der genutzten Fooxx-Nutzer des Tests

So hat z.B. der Nutzer VAH mit den Nutzer ROADRUNNERLENNY einen Ähnlichkeitswert von 0,00233645. Die Berechnung und Normalisierung der Ähnlichkeit ist bereits im bestehenden Fooxx implementiert und wird für diesen Test abgefragt. Da JKESSLER ein seltener Fooxx-Nutzer ist, besitzt er keine Ähnlichkeit mit den anderen Nutzern. Wie sich auch im Laufe des Tests herausstellen wird, ist für ihn somit keine Personalisierung möglich. Die Ähnlichkeitswerte werden im Laufe des Tests mit berücksichtigt.

Nach Durchführung des Tests mit den Testern wurden die gesammelten Informationen der Fooxx-Toolbar dazu genutzt, eine Relevanzbewertung der aufgerufenen Webseiten anhand des vorgestellten personalisierten Sortierungsalgorithmus mit Hilfe der Testumgebung vorzunehmen. Diese automatisierte Bewertung wurde dann mit der von den Testern erstellten Relevanzbewertung verglichen. Es wird analysiert, ob der hier vorgestellte Algorithmus sinnvolle Ergebnisse liefert.

6.1.1 Testergebnisse

Nach Durchführung des Tests wurde eine Liste erstellt, auf denen die von den Testern als relevant eingestuften Seiten für die zwei Aufgaben des Aufgabenblatts aufgelistet werden (Anhang B2 und B3). Die relevanten Seiten wurden dann zur besseren Übersicht nach Ihrer Häufigkeit sortiert und in verschiedene Plätze gruppiert. Diese Auswertung wurde für alle Tester erstellt (Anhang B4 und B9), sowie

einmal getrennt für die einzelnen Fooxx-Nutzer, die jeweils von den Testern benutzt wurden. (Anhang B5 und B10 für Nutzer DASDINGSDA, Anhang B6 und B11 für den Nutzer VAH, Anhang B7 und B12 für Nutzer ROADRUNNERLENNY und Anhang B8 und B13 für den Nutzer JKESSLER.)

Um die Ergebnisse des Personalisierungsalgorithmus zu testen, wurde die Testumgebung genutzt. Zunächst wurden die unpersonalisierten Ergebnisse der Suchmaschinen Google und Fooxx mit der Abfrage „Vorzüge von Tee“ und einer Maximalmenge von 100 Treffern abgefragt. Für die Abfrage von Fooxx wurde dabei als Nutzer „GAST“ angegeben. GAST ist ein Nutzer, der von jedem verwendet werden kann, und somit ein sehr allgemeines Profil besitzt. Er stellt somit einen universellen Nutzer dar und liefert nur eine geringe Personalisierung der Ergebnisliste. Danach wurde der Suchbegriff „Abholzung Regenwald“ benutzt. Da Fooxx als Meta-Suchmaschine auch Google abfragt, sind Überschneidungen der Ergebnisse vorhanden. Auf die Abfrage des eigenen Index wurde verzichtet, da die hier ausgegebene Ergebnisliste einen sehr großen Teil an nicht relevanten Seiten bzw. an Seiten zurücklieferte, die nicht in den von den Testern als relevant bewerteten Seiten vorkamen. Eine Einbeziehung des eigenen Index wäre hier nicht sinnvoll gewesen, da eine Vergleichbarkeit mit den Testresultaten nicht möglich wäre. Das der eigene Index solch eine schlechte verwendbare Ergebnisliste zurücklieferte, ist auf die mangelnde Leistungsfähigkeit und die nur testweise für diese Arbeit vorgenommene Implementierung der Volltextsuche Lucene zum Durchsuchen des Index zurückzuführen.

Im Folgenden werden Auszüge der unpersonalisierten Ergebnislisten abgebildet und besprochen. In diesen Auszügen sind nur die Treffer aufgeführt, die von den Testern auch als relevant eingestuft wurden. Dort wird die URL des Treffers, sein Rang in der jeweiligen Suchmaschine sowie der Platz in der Rangfolge der relevanten Seiten angegeben. Die kompletten unpersonalisierten Ergebnislisten sind in den Tabellen im Anhang B14, B15, B16 und B17 der Vollständigkeit halber aufgeführt.

Anschließend wurden die beiden Ergebnislisten mit dem im Kapitel 4.7 beschriebenen Personalisierungs-Algorithmus g_1 neu sortiert. Als Faktor k wurde der Wert 2 gewählt, da bei diesem Wert die sinnvollste Personalisierung

anzunehmen ist. Dies wird in 6.2.2 näher erläutert. Als Nutzer für die Personalisierung wurden die Fooxx-Nutzer DASDINGSDA, VAH, ROADRUNNERLENNY und JKESSLER mit Ihrer entsprechenden ID angegeben. Dann wurde die Personalisierung durchgeführt. Für beide Suchbegriffe und jeweils beide Suchmaschinen sind die personalisierten Ergebnislisten ermittelt worden. Für die jeweils ersten 15 Treffer der personalisierten Ergebnislisten wurden die jeweiligen Plätze in der Relevanzbewertung der Tester ermittelt. Der Rang vor der Sortierung ist jeweils in Klammern angegeben. Des Weiteren wird angegeben, ob ein beschreibender Wert für diesen Treffer existierte. Der Wert selbst spielt dabei eine untergeordnete Rolle, da alle beschreibenden Werte aufgrund der großen Anzahl an Interaktionen im System sehr klein sind und somit wenig Aussagekraft haben. Im Folgenden werden diese personalisierten Ergebnislisten nun einzeln vorgestellt, wobei nur die ersten 15 Treffer der Ergebnisliste berücksichtigt werden. Die weiter hinten liegenden Treffer besitzen keine Aussagekraft für diesen Test. Sie sind der Vollständigkeit halber im Anhang B18, B19, B20 und B21 aufgeführt.

6.1.2 Auswertung der Ergebnisse

Es werden die Ergebnislisten vor und nach der Personalisierung für die einzelnen Nutzer betrachtet. Bei den unpersonalisierten Ergebnislisten werden nur die Seiten dargestellt, die auch von den Testern als relevant eingestuft wurden. Bei den personalisierten Ergebnislisten werden nur die ersten 15 Ergebnisse mit einbezogen, da bei weiter hinten liegenden Treffern meist kein beschreibender Wert mehr ermittelt werden kann. Zudem ist für einen Nutzer in der Regel nur die erste Trefferseite interessant. Meist wird eher der Suchbegriff geändert, als die folgenden Trefferseiten zu besichtigen. Es werden nun zunächst Auszüge der unpersonalisierten Trefferlisten in den Tabellen 3,4,5 und 6 abgebildet. Diese werden im weiteren Verlauf dieses Kapitels noch eingehender erläutert.

------ Ergebnisliste Google für den Suchbegriff "Vorzüge von Tee": -------	Platz b. Testern
1 : http://www.wander.ch/cgi/de/magazine/contact_archive/contact_21/contact1.asp	8
5 : http://teeverband.at/wiss-sort01.htm	1
17 : http://www.swr.de/kaffee-oder-tee/essen/tee/2003/04/16	7
19 : http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	6
23 : http://www.teebuch.de/005008.htm	7
30 : http://www.abnehmtreff.de/article133.html	7
34 : http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04	2
44 : http://www.abnehmtreff.de/article133.html	7
51 : http://de.isodisnatura.ch/nutrition_-_article.htm?ID=14	7
52 : http://www.neuesleben.net/archive/text/gruentee01.shtml	7
60 : http://www.kapstadt-news.de/news/283.htm	8
99 : http://deutschesfachbuch.de/info/detail.php?isbn=3453141784	7

Tabelle 3: Unpersonalisierte Ergebnisliste von Google für den Suchbegriff „Vorzüge von Tee“

------ Ergebnisliste Fooxx für den Suchbegriff "Vorzüge von Tee": ------	Platz b. Testern
1 : http://www.tee-special.com/tee.html	1
2 : http://www.pandorahs-traum.de/schatztruhe/tee.html	7
3 : http://www.wendtour.de/ernaehrungtee.htm	8
4 : http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	4
5 : http://www.frauke-spehr.info	6
6 : http://www.steffys.de/htm/gruenertee.htm	8
7 : http://www.teebuch.de/teebuch.txt	5
8 : http://teeverband.at/wiss-sort01.htm	1
9 : http://www.pralinen-teeke.de/shop/html/tee.html	8
10 : http://www.teebuch.de/005008.htm	7
11 : http://www.teeimport.com/teeabc.htm	8
12 : http://www.neuesleben.net/archive/text/gruentee01.shtml	7
14 : http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	3
15 : http://www.geocities.com/Athens/Delphi/3967/tee.html	8
49 : http://www.dreisesselapotheke.de/htm/ernaehrung/99_10-61_7548_TeeZeit_2_2003.pdf	7

Tabelle 4: Unpersonalisierte Ergebnisliste von Fooxx für den Suchbegriff „Vorzüge von Tee“

------ Ergebnisliste Google für den Suchbegriff "Abholzung Regenwald": -------	Platz b. Testern
1 : http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	1
2 : http://www.regenwald.org/new/amazonas/highnoon.htm	1
5 : http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	4
8 : http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	2
10 : http://www.regenwald-spende.de/ueber_uns.htm	2
11 : http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	2
17 : http://www.der-gruene-faden.de/text/text648.html	2
20 : http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	4
21 : http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms	3
40 : http://www.umg.at/112001/abholzung.php	4
79 : http://www.regenwald.org/new/regenwaldreport/artikel.php?id=147	5
83 : http://berg.heim.at/tibet/450508/Regen.htm	4
90 : http://www.econautix.de/site/econautixpage_46.php	3

Tabelle 5: Unpersonalisierte Ergebnisliste von Google für den Suchbegriff „Abholzung Regenwald“

------ Ergebnisliste Fooxx für den Suchbegriff "Abholzung Regenwald": -------	Platz b. Testern
9 : http://www.regenwald.org/new/amazonas/highnoon.htm	1
10 : http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	4
12 : http://www.der-gruene-faden.de/text/text648.html	2
13 : http://www.econautix.de/site/econautixpage_46.php	3
15 : http://www.regenwald-spende.de/ueber_uns.htm	2
18 : http://www.regenwaldschutz.de	5
20 : http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	4
23 : http://www.umg.at/112001/abholzung.php	4
26 : http://www.regenwald.org/new/regenwaldreport/artikel.php?id=147	5
29 : http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	2
32 : http://www.cil-frankfurt.de/programme/NEINN/MixMax/Dokumente/CRRegenwald.htm	5
42 : http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms	3
63 : http://regenwald-spende.de/ueber_uns.htm	2

Tabelle 6: Unpersonalisierte Ergebnisliste von Fooxx für den Suchbegriff „Abholzung Regenwald“

Betrachtet man die Auszüge der unpersonalisierten Ergebnislisten (Tab. 3,4,5 und 6), wird man feststellen, dass sich meist weniger als die Hälfte der von den Testern als relevant bewerteten Seiten überhaupt in den Ergebnislisten findet. Obwohl die Abfrage und die Suchmaschinen gleich geblieben sind, liefern Suchmaschinen selten gleiche Ergebnislisten. Dies hängt z.B. bei Google von dem Rechenzentrum ab, das für die Anfrage einbezogen wird – die verwalteten Indexe in den Rechenzentren arbeiten nicht synchron zu anderen Rechenzentren und somit treten Abweichungen auf. Auch wurde die Auswertung des Tests zeitlich versetzt zum Test selber durchgeführt, welches bei dem schnelllebigen Wandel im Internet ebenfalls Änderungen in den Indexen der Suchmaschinen und somit Änderungen in den Ergebnislisten bewirkte.

6.1.2.1 Auswertung Suchbegriff „Vorzüge von Tee“ in Google

Bei der Betrachtung der unpersonalisierten Ergebnisliste von Google (Tab. 3) kann man erkennen, dass sich die elf von den Testern als relevant bewerteten Seiten größtenteils gleichmäßig über die ersten 60 Ergebnisse verteilen. Um somit den Großteil der von den Testern als relevant bewerteten Seiten angezeigt zu bekommen, müsste ein Nutzer bei Google fünfmal weiterblättern. Eine große Menge an für die Tester nicht relevanten Seiten befindet sich somit in der Ergebnisliste.

Nach der Personalisierung der Ergebnisliste von Google für den Nutzer DASDINGSDA ergibt sich folgende Tabelle (Tab. 7). Der Rang in Google vor der Sortierung ist jeweils in Klammern angegeben. Weiterhin ist der Platz bei der Relevanzbewertung der Tester aufgeführt und angegeben, ob ein beschreibender Wert für das Dokument existierte.

Nutzer DASDINGSDA - Personalisierte Ergebnisliste Google	Platz	B. Wert
1 (40) : http://www.bluegreen.net/deutsch/info/texte/stevia.htm	N/A	Ja
2 (5) : http://teeverband.at/wiss-sort01.htm	1	Ja
3 (34) : http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04	2	Ja
4 (1) : http://www.wander.ch/cgi/de/magazine/contact_archive/contact_21/contact1.asp	8	Ja
5 (52) : http://www.neuesleben.net/archive/text/gruentee01.shtml	7	Ja
6 (99) : http://deutschesfachbuch.de/info/detail.php?isbn=3453141784	7	Ja
7 (23) : http://www.teebuch.de/005008.htm	7	Ja
8 (19) : http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	6	Ja
9 (8) : http://www.milando.de/sq+tee+_167.htm	N/A	Nein
10 (9) : https://grubauer.de/product_info.php?products_id=17	N/A	Nein
11 (11) : http://www.heuschrecke.com	N/A	Nein
12 (10) : http://www.fm-online.at/jaos/page/main_heute.tmpl?article_id=10010862	N/A	Nein
13 (7) : http://www.tee-seite.de/aassa.htm	N/A	Nein
14 (14) : http://asconet.org:8000/antville/labor	N/A	Nein
15 (15) : http://www.preisglocke.de/shop/i_m_naturkosmetik_duschbaeder_+_lotionen_gruener_	N/A	Nein

Tabelle 7 : Personalisierte Ergebnisliste des Nutzers DASDINGSDA der Suchmaschine Google

Zunächst fällt bei Betrachtung von Tabelle 7 auf, dass sich nun unter den ersten zehn Treffern sieben relevante Seiten befinden. Diese Seiten sind meist von den hinteren Plätzen hoch gestuft worden.

Auf Rang 1 der personalisierten Bewertung liegt eine Seite, welche sich von Rang 40 auf Rang 1 „hochgearbeitet“ hat. Diese Seite wurde allerdings von keinem Tester als relevant eingestuft. Diese Veränderung ist auf die schon vorhandenen Daten des bestehenden Fooxx-Nutzer DASDINGSDA zurückzuführen. Auf Rang 2 liegt eine Seite, die vom Google-Rang 5 hoch gestuft wurde und von den Testern als mit die relevanteste Seite bewertet wurde. Der 1. Rang von Google liegt nach der Personalisierung auf Rang 4 und ist eine Seite, die Platz 7 der Relevanzbewertung belegt. Die Ränge 3,5,6,7 und 8 sind auffällige Merkmale für eine Neusortierung. Die vorherigen Google-Ränge 34, 5, 99, 23 und 19 sind verhältnismäßig weit oben eingeordnet worden. Sie alle besitzen eine gewisse Relevanz für verschiedene Testpersonen.

Ab Rang 9 liegt kein beschreibender Wert mehr für die einzelnen Seiten vor und es sind keine weiteren Seiten vorhanden, die vom Algorithmus bewertet werden könnten. Die Sortierung ab diesem Platz unterliegt nicht mehr der Sortierung des Algorithmus. Auftretende Sortierungen weiter hinten in der Liste sind abhängig von der Implementierung der Sortierung von Seiten ohne beschreibenden Wert. Seiten ohne beschreibenden Wert unterliegen dabei keiner besonderen Sortierung; bei Ihnen kann somit keine besondere Ordnung ausgemacht werden.

Nach der Personalisierung der Ergebnisliste von Google für den Nutzer ROADRUNNERLENNY erhält man Tabelle 8.

Nutzer ROADRUNNERLENNY - Personalisierte Ergebnisliste Google	Platz	B. Wert
1 (5) : http://teeverband.at/wiss-sort01.htm	1	Ja
2 (34) : http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04	2	Ja
3 (23) : http://www.teebuch.de/005008.htm	7	Ja
4 (99) : http://deutschesfachbuch.de/info/detail.php?isbn=3453141784	7	Ja
5 (1) : http://www.wander.ch/cgi/de/magazine/contact_archive/contact_21/contact1.asp	8	Ja
6 (52) : http://www.neuesleben.net/archive/text/gruentee01.shtml	7	Ja
7 (30) : http://www.abnehmtreff.de/article133.html	7	Ja
8 (19) : http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	6	Ja
9 (33) : http://teewalter.de/info35.php	N/A	Ja
10 (51) : http://de.isodisnatura.ch/nutrition_-_article.htm?ID=14	7	Ja
11 (44) : http://www.abnehmtreff.de/article133.html	7	Ja
12 (60) : http://www.kapstadt-news.de/news/283.htm	8	Ja
13 (40) : http://www.bluegreen.net/deutsch/info/texte/stevia.htm	N/A	Ja
14 (13) : http://www.bauarchiv.de/shopzone24/product.php?asin=B0007WDHZQ&section=4&mc	N/A	Nein
15 (14) : http://asconet.org:8000/antville/labor	N/A	Nein

Tabelle 8: Personalisierte Ergebnisliste des Nutzers ROADRUNNERLENNY der Suchmaschine Google

Bei der Personalisierung für den Nutzer ROADRUNNERLENNY schaffen es sogar alle elf relevanten Seiten unter die ersten zwölf. Ab Treffer 14 existiert kein weiterer beschreibender Wert mehr. Es sind Überschneidungen der ersten 8 Treffer mit der Personalisierung des Nutzers DASDINGSDA in der Reihenfolge erkennbar. (Tab. 8)

Nutzer VAH - Personalisierte Ergebnisliste Google	Platz	B. Wert
1 (5) : http://teeverband.at/wiss-sort01.htm	1	Ja
2 (34) : http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04	2	Ja
3 (19) : http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	6	Ja
4 (60) : http://www.kapstadt-news.de/news/283.htm	8	Ja
5 (44) : http://www.abnehmtreff.de/article133.html	7	Ja
6 (51) : http://de.isodisnatura.ch/nutrition_-_article.htm?ID=14	7	Ja
7 (33) : http://teewalter.de/info35.php	N/A	Ja
8 (30) : http://www.abnehmtreff.de/article133.html	7	Ja
9 (1) : http://www.wander.ch/cgi/de/magazine/contact_archive/contact_21/contact1.asp	8	Ja
10 (40) : http://www.bluegreen.net/deutsch/info/texte/stevia.htm	N/A	Ja
11 (10) : http://www.fm-online.at/jaos/page/main_heute.tmpl?article_id=10010862	N/A	Nein
12 (9) : https://grubauer.de/product_info.php?products_id=17	N/A	Nein
13 (8) : http://www.milando.de/sq+tee+_167.htm	N/A	Nein
14 (7) : http://www.tee-seite.de/aassa.htm	N/A	Nein
15 (6) : http://de.chinabroadcast.cn/other/teekultur/23.htm	N/A	Nein

Tabelle 9: Personalisierte Ergebnisliste des Nutzers VAH der Suchmaschine Google

Bei der Personalisierung des Nutzers VAH sind acht der als relevant bewerteten Seiten unter den ersten zehn Treffern. Die Ähnlichkeit der Reihenfolge der Ergebnisse zu den Nutzern DASDINGSDA (Tab. 7) und ROADRUNNERLENNY (Tab. 8) ist eher gering. Betrachtet man aber die Plätze der Relevanzbewertung der Test, so ist hier ebenfalls eine Ähnlichkeit erkennbar.

Nutzer JKESSLER - Personalisierte Ergebnisliste Google für Aufgabe 1	Platz	B. Wert
1 (1) : http://www.wander.ch/cgi/de/magazine/contact_archive/contact_21/contact1.asp	8	Nein
2 (2) : http://www.wander.ch/cgi/de/magazine/contact_current/contact1.asp?print=yes	N/A	Nein
3 (3) : http://www.miele.de/D/PR_kva_04.html	N/A	Nein
4 (4) : http://www.das-grosse-leben.de/htm/aus%20aller%20Welt.html	N/A	Nein
5 (5) : http://teeverband.at/wiss-sort01.htm	1	Nein
6 (6) : http://de.chinabroadcast.cn/other/teekultur/23.htm	N/A	Nein
7 (7) : http://www.tee-seite.de/aassa.htm	N/A	Nein
8 (8) : http://www.milando.de/sq+tee+_167.htm	N/A	Nein
9 (9) : https://grubauer.de/product_info.php?products_id=17	N/A	Nein
10 (10) : http://www.fm-online.at/jaos/page/main_heute.tmpl?article_id=10010862	N/A	Nein
11 (11) : http://www.heuschrecke.com	N/A	Nein
12 (12) : http://www.dooyoo.de/wasserfilter/brita-fjord/1021751	N/A	Nein
13 (13) : http://www.bauarchiv.de/shopzone24/product.php?asin=B0007WDHZQ&section=4&r	N/A	Nein
14 (14) : http://asconet.org:8000/antville/labor	N/A	Nein
15 (15) : http://www.preisglocke.de/shop/i_m_naturkosmetik_duschbaeder_+_lotionen_gruene	N/A	Nein

Tabelle 10:Personalisierte Ergebnisliste des Nutzers JKESSLER der Suchmaschine Google

Bei der Personalisierung des Nutzers JKESSLER (Tab. 10) ist erkennbar, dass keine Veränderung der Trefferliste vorgenommen und somit kein beschreibender Wert errechnet wurde. Dies ist auf die geringe Anzahl an Interaktionen zurückzuführen, die der Nutzer JKESSLER bis jetzt bei Fooxx aufgrund seiner seltenen Nutzung erzeugt hatte. Es kann somit mit dem Algorithmus wegen der fehlenden Ähnlichkeitswerte keine Personalisierung vorgenommen werden. Die personalisierten Ergebnislisten dieses Nutzers entfallen somit aus der weiteren Betrachtung.

6.1.2.2 Auswertung Suchbegriff „Vorzüge von Tee“ in Fooxx

Bei der Betrachtung der unpersonalisierten Ergebnisse von Fooxx (Tab. 4) ist erkennbar, dass sich 14 der 15 von den Testern als relevant bewerteten Seiten untern den ersten 15 Treffern befinden. Dies sieht zunächst nach einem sehr guten Suchergebnis aus, ist aber auf die bereits integrierte Personalisierung von Fooxx zurückzuführen: Da während des Test die Tester Interaktionen erzeugt haben, wurde bereits eine Bewertung der Seite mit Hilfe des bereits erklärten „inneren Wert“ vorgenommen. Dadurch haben die von den Testern als relevant eingestuften Seiten schon während des Tests einen hohen inneren Wert zugewiesen bekommen und tauchen dementsprechend in der Ergebnisliste von Fooxx weit oben auf.

Nutzer DASDINGSDA - Ergebnisliste Fooxx	Platz	B. Wert
1 (40) : http://www.makrobiotik.com/lexikon.htm	N/A	Ja
2 (7) : http://www.teebuch.de/teebuch.txt	5	Ja
3 (8) : http://teeverband.at/wiss-sort01.htm	1	Ja
4 (2) : http://www.pandorahs-traum.de/schatztruhe/tee.html	7	Ja
5 (3) : http://www.wendtour.de/ernaehrungtee.htm	8	Ja
6 (1) : http://www.tee-special.com/tee.html	1	Ja
7 (10) : http://www.teebuch.de/005008.htm	7	Ja
8 (12) : http://www.neuesleben.net/archive/text/gruentee01.shtml	7	Ja
9 (11) : http://www.teeimport.com/teeabc.htm	8	Ja
10 (5) : http://www.frauke-spehr.info	6	Ja
11 (4) : http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	4	Ja
12 (9) : http://www.pralinen-teeke.de/shop/html/tee.html	8	Ja
13 (6) : http://www.steffys.de/htm/gruenertee.htm	8	Ja
14 (15) : http://www.geocities.com/Athens/Delphi/3967/tee.html	8	Ja
15 (14) : http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	3	Nein

Tabelle 11: Personalisierte Ergebnisliste des Nutzers DASDINGSDA der Suchmaschine Fooxx

Betrachtet man die personalisierten Ergebnislisten des Nutzers DASDINGSDA (Tab. 11), so zeigt sich, dass sich auf Platz 1 eine Seite befindet, die nicht in der Relevanzbewertung der Tester auftaucht. Dies ist wiederum mit dem bestehenden Profil des Nutzers zu erklären. Für die ersten 14 Treffer wurde ein beschreibender Wert ermittelt. Bei der Sortierung dieser Seiten ist auffällig, dass eine Umsortierung der Reihenfolge vorgenommen wurde – aber eine bestimmte Sortierung, z.B. anhand der Relevanzbewertung der Tester sortiert nach den Nutzern, ist nicht erkennbar.

Ebenso verhält es sich mit den Ergebnislisten der Nutzer ROADRUNNERLENNY (Tab. 12) und VAH (Tab. 13). Auch hier können keine konkreten Aussagen über die neue Sortierung gemacht werden.

Nutzer ROADRUNNERLENNY - Ergebnisliste Fooxx	Platz	B. Wert
1 (8) : http://teeverband.at/wiss-sort01.htm	1	Ja
2 (7) : http://www.teebuch.de/teebuch.txt	5	Ja
3 (3) : http://www.wendtour.de/ernaehrungtee.htm	8	Ja
4 (1) : http://www.tee-special.com/tee.html	1	Ja
5 (2) : http://www.pandorahs-traum.de/schatztruhe/tee.html	7	Ja
6 (10) : http://www.teebuch.de/005008.htm	7	Ja
7 (12) : http://www.neuesleben.net/archive/text/gruentee01.shtml	7	Ja
8 (11) : http://www.teeimport.com/teeabc.htm	8	Ja
9 (9) : http://www.pralinen-teeke.de/shop/html/tee.html	8	Ja
10 (4) : http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	4	Ja
11 (5) : http://www.frauke-spehr.info	6	Ja
12 (6) : http://www.steffys.de/htm/gruenertee.htm	8	Ja
13 (49) : http://www.dreisesselapotheke.de/htm/ernaehrung/99_10-61_7548_TeeZeit_2_2003.pd	7	Ja
14 (40) : http://www.makrobiotik.com/lexikon.htm	N/A	Ja
15 (14) : http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	3	Nein

Tabelle 12: Personalisierte Ergebnisliste des Nutzers ROADRUNNERLENNY der Suchmaschine Fooxx

Bei der Trefferliste für den Nutzer ROADRUNNERLENNY in Tabelle 12 wurde ebenfalls für die ersten 14 Treffer ein beschreibender Wert ermittelt. Die Sortierung lässt keine Interpretation zu.

Nutzer VAH - Ergebnisliste Fooxx	Platz	B. Wert
1 (7) : http://www.teebuch.de/teebuch.txt	5	Ja
2 (8) : http://teeverband.at/wiss-sort01.htm	1	Ja
3 (3) : http://www.wendtour.de/ernaehrungtee.htm	8	Ja
4 (2) : http://www.pandorahs-traum.de/schatztruhe/tee.html	7	Ja
5 (1) : http://www.tee-special.com/tee.html	1	Ja
6 (49) : http://www.dreisesselapotheke.de/htm/ernaehrung/99_10-61_7548_TeeZeit_2_2003.pdf	7	Ja
7 (15) : http://www.geocities.com/Athens/Delphi/3967/tee.html	8	Ja
8 (40) : http://www.makrobiotik.com/lexikon.htm	N/A	Nein
9 (5) : http://www.frauke-spehr.info	6	Nein
10 (4) : http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	4	Nein
11 (11) : http://www.teeimport.com/teeabc.htm	8	Nein
12 (12) : http://www.neuesleben.net/archive/text/gruentee01.shtml	7	Nein
13 (6) : http://www.steffys.de/htm/gruenertee.htm	8	Nein
14 (14) : http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	3	Nein
15 (9) : http://www.pralinen-teeke.de/shop/html/tee.html	8	Nein

Tabelle 13: Personalisierte Ergebnisliste des Nutzers VAH der Suchmaschine Fooxx

Für die Sortierung des Nutzers VAH in Tabelle 13 konnte nur für die ersten acht Treffer ein beschreibender Wert ermittelt werden. Die restlichen Treffer sind somit für die Sortierung uninteressant. Auch hier kann keine besondere Aussage über die Sortierung gemacht werden.

6.1.2.3 Auswertung Suchbegriff „Abholzung Regenwald“ in Google

In der unpersonalisierten Ergebnisliste von Google für den Suchbegriff „Abholzung Regenwald“ (Tab. 5) finden sich 13 der insgesamt 22 von den Testern als relevant bewerteten Seiten in der Ergebnisliste. Neun davon befinden sich unter den ersten 21 Treffern. Die restlichen vier liegen großteils im hinteren Drittel der Ergebnisse.

------ Ergebnisse Personalisierung Google für Nutzer DASDINGSDA: -------	Platz	B. Wert
1 (10) : http://www.regenwald-spende.de/ueber_uns.htm	2	Ja
2 (9) : http://www.referate-seite.com/referatzimmermann.html	N/A	Ja
3 (8) : http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	2	Ja
4 (17) : http://www.der-gruene-faden.de/text/text648.html	2	Ja
5 (21) : http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms	3	Ja
6 (11) : http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	2	Ja
7 (5) : http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	4	Ja
8 (2) : http://www.regenwald.org/new/amazonas/highnoon.htm	1	Ja
9 (1) : http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	1	Ja
10 (30) : http://www.4teachers.de/material/2656/Abholzung_im_tropischen_Regenwald.html	N/A	Ja
11 (40) : http://www.umg.at/112001/abholzung.php	4	Ja
12 (20) : http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	5	Ja
13 (87) : http://www.wuestenfahrer.de/westafrika/westafrika_regenwald.htm	N/A	Ja
14 (83) : http://berg.heim.at/tibet/450508/Regen.htm	4	Ja
15 (90) : http://www.econautix.de/site/econautixpage_46.php	3	Ja

Tabelle 14: Personalisierte Ergebnisliste des Nutzers DASDINGSDA der Suchmaschine Google

Betrachtet man die personalisierten Ergebnisse der Suchmaschine Google für den Nutzer DASDINGSDA (Tab. 14), so ist eine Verdichtung der relevanten Ergebnisse auf die ersten 15 Treffer zu verzeichnen. 13 der 14 relevanten Treffer liegen nun unter den ersten 15, lediglich ein weiterer relevanter Treffer liegt in der Ergebnisliste weiter hinten. Besonders auffällig ist, dass die Treffer auf Rang 14 und 15 bei Google auf Rang 83 und 90 lagen. Weiterhin wurden die von Google auf Rang 1 und 2 platzierten und ebenfalls von den Testern als mit am relevantesten eingestuften Seiten auf die Ränge 9 und 10 gelegt.

Eine besondere Sortierung dieser Treffer in Bezug auf die Personalisierung ist im Vergleich zu der Relevanzeinordnung der Tester, sortiert nach dem jeweiligen Fooxx-Nutzer, nicht zu erkennen.

------ Ergebnisse Personalisierung Google für Nutzer ROADRUNNERLENNY: -------	Platz	B. Wert
1 (5) : http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	4	Ja
2 (8) : http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	2	Ja
3 (10) : http://www.regenwald-spende.de/ueber_uns.htm	2	Ja
4 (17) : http://www.der-gruene-faden.de/text/text648.html	2	Ja
5 (2) : http://www.regenwald.org/new/amazonas/highnoon.htm	1	Ja
6 (1) : http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	1	Ja
7 (11) : http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	2	Ja
8 (21) : http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms	3	Ja
9 (30) : http://www.4teachers.de/material/2656/Abholzung_im_tropischen_Regenwald.html	N/A	Ja
10 (9) : http://www.referate-seite.com/referatzimmermann.html	N/A	Ja
11 (90) : http://www.econautix.de/site/econautixpage_46.php	3	Ja
12 (40) : http://www.umg.at/112001/abholzung.php	4	Ja
13 (7) : http://www.energieportal24.de/abholzung%20regenwald/qry_abholzung%20regenwal	N/A	Nein
14 (14) : http://www.referate-seite.com/referate-von-sportstudios.html	N/A	Nein
15 (12) : http://www.referate-seite.com/referat--weisse-rose.html	N/A	Nein

Tabelle 15: Personalisierte Ergebnisse des Nutzers ROADRUNNERLENNY der Suchmaschine Google

Bei der Personalisierung der Ergebnisse von Google für den Nutzer ROADRUNNERLENNY (Tab. 15) sind zehn der relevanten Seiten unter die ersten zwölf Ergebnisse gekommen. Danach gab es keine weiteren Seiten mit einem beschreibenden Wert mehr.

Rang elf und zwölf zeigen dabei, dass die Google-Treffer von Rang 90 bzw. 40 jetzt wesentlich weiter vorne liegen. Insgesamt ist wieder eine Verdichtung der relevanten Treffer im vorderen Bereich zu beobachten. Die beiden Top-Platzierungen von Google und von den Testern liegen nach der Personalisierung auf Rang fünf und sechs.

------ Ergebnisse Personalisierung Google für Nutzer VAH: -------	Platz	B. Wert
1 (10) : http://www.regenwald-spende.de/ueber_uns.htm	2	Ja
2 (8) : http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	2	Ja
3 (7) : http://www.energieportal24.de/abholzung%20regenwald/qry_abholzung%20regenwal	N/A	Ja
4 (90) : http://www.econautix.de/site/econautixpage_46.php	3	Ja
5 (40) : http://www.umg.at/112001/abholzung.php	4	Ja
6 (17) : http://www.der-gruene-faden.de/text/text648.html	2	Ja
7 (2) : http://www.regenwald.org/new/amazonas/highnoon.htm	1	Ja
8 (1) : http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	1	Ja
9 (5) : http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	4	Ja
10 (87) : http://www.wuestenfahrer.de/westafrika/westafrika_regenwald.htm	N/A	Ja
11 (20) : http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	4	Ja
12 (21) : http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms	3	Ja
13 (83) : http://berg.heim.at/tibet/450508/Regen.htm	4	Ja
14 (82) : http://www.abenteuer-regenwald.de/tiere.php	N/A	Ja
15 (79) : http://www.regenwald.org/new/regenwaldreport/artikel.php?id=147	5	Ja

Tabelle 16: Personalisierte Ergebnisliste des Nutzers VAH der Suchmaschine Google

Für den Nutzer VAH sind ähnliche Beobachtungen wie für den Nutzer DASDINGSDA und ROADRUNNERLENNY zu machen (Tab. 16). Eine Verdichtung der relevanten Seiten unter den ersten 15 Treffern ist zu erkennen. So waren Rang 4 und 5 vorher bei Google auf Rang 90 und 40 gelistet. Die von Google und von den Testern als sehr relevant eingestuften Seiten liegen nun auf Rang 7 und 8. Eine besondere Sortierung in diesem Bereich, die auf eine Personalisierung der weiter vorne liegenden Treffer hindeutet, ist nicht zu erkennen.

6.1.2.4 Auswertung Suchbegriff „Abholzung Regenwald“ in Fooxx

Die unpersonalisierte Ergebnisliste von Fooxx (Tab. 6) zeigt eine Ansammlung von 13 relevanten Treffern größtenteils im vorderen Drittel der Ergebnisliste.

------ Ergebnisse Personalisierung Fooxx für Nutzer DASDINGSDA: -------	Platz	B.Wert
1 (9) : http://www.regenwald.org/new/amazonas/highnoon.htm	1	Ja
2 (15) : http://www.regenwald-spende.de/ueber_uns.htm	2	Ja
3 (10) : http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	4	Ja
4 (12) : http://www.der-gruene-faden.de/text/text648.html	2	Ja
5 (11) : http://www.referate-seite.com/referatzimmermann.html	N/A	Ja
6 (42) : http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms	3	Ja
7 (46) : http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	2	Ja
8 (29) : http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	2	Ja
9 (45) : http://www.4teachers.de/material/2656/Abholzung_im_tropischen_Regenwald.html	N/A	Ja
10 (20) : http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	4	Ja
11 (23) : http://www.umg.at/112001/abholzung.php	4	Ja
12 (22) : http://www.wuestenfahrer.de/westafrika/westafrika_regenwald.htm	N/A	Ja
13 (54) : http://berg.heim.at/tibet/450508/Regen.htm	4	Ja
14 (56) : http://www.regenwald-institut.de/deutsch/Aktionen/Bolivien.htm	5	Ja
15 (25) : http://www.faszination-regenwald.de/info-center/oekosystem/wasserhaushalt.htm	N/A	Ja

Tabelle 17: Personalisierte Ergebnisliste des Nutzers DASDINGSA der Suchmaschine Fooxx

Nach der Personalisierung für den Nutzer DASDINGSDA befinden sich 11 relevante Seiten unter den ersten 15 Treffern (Tab. 17). Dabei ist die Sortierung der Plätze der relevanten Treffer recht nah an der gewünschten Reihenfolge. Die hinteren relevanten Seiten sind somit wie gewünscht in den vorderen Bereich gewandert.

------ Ergebnisse Personalisierung Fooxx für Nutzer ROADRUNNERLENNY: -------	Platz	B.Wert
1 (9) : http://www.regenwald.org/new/amazonas/highnoon.htm	1	Ja
2 (15) : http://www.regenwald-spende.de/ueber_uns.htm	2	Ja
3 (12) : http://www.der-gruene-faden.de/text/text648.html	2	Ja
4 (10) : http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	4	Ja
5 (29) : http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	2	Ja
6 (45) : http://www.4teachers.de/material/2656/Abholzung_im_tropischen_Regenwald.html	N/A	Ja
7 (11) : http://www.referate-seite.com/referatzimmermann.html	N/A	Ja
8 (46) : http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	2	Ja
9 (42) : http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms	3	Ja
10 (18) : http://www.regenwaldschutz.de	5	Ja
11 (19) : http://www.waldportal.org/books.taiga/print.html	N/A	Ja
12 (21) : http://www.brasilien.de/land/florafauna/tropregenwald.asp	N/A	Ja
13 (78) : http://www.umweltbrief.de/neu/html/archiv/Regenwald.txt	N/A	Ja
14 (32) : http://www.cil-frankfurt.de/programme/NEINN/MixMax/Dokumente/CRRegenwald.htm	5	Ja
15 (23) : http://www.umg.at/112001/abholzung.php	4	Ja

Tabelle 18: Personalisierte Ergebnisliste des Nutzers ROADRUNNERLENNY der Suchmaschine Fooxx

Die Personalisierung von Fooxx für den Nutzer ROADRUNNERLENNY zeigt wieder eine Verdichtung der relevanten Seiten unter den ersten 15 Treffern (Tab. 18). Eine besondere Reihenfolge bzw. Personalisierung kann man nicht erkennen.

------ Ergebnisse Personalisierung Fooxx für Nutzer VAH: -------	Platz	B.Wert
1 (23) : http://www.umg.at/112001/abholzung.php	4	Ja
2 (29) : http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	2	Ja
3 (9) : http://www.regenwald.org/new/amazonas/highnoon.htm	1	Ja
4 (10) : http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	4	Ja
5 (15) : http://www.regenwald-spende.de/ueber_uns.htm	2	Ja
6 (13) : http://www.econautix.de/site/econautixpage_46.php	3	Ja
7 (12) : http://www.der-gruene-faden.de/text/text648.html	2	Ja
8 (24) : http://gruppen.greenpeace.de/koblenz/ini2000plus.htm	N/A	Ja
9 (32) : http://www.cil-frankfurt.de/programme/NEINN/MixMax/Dokumente/CRRegenwald.htm	5	Ja
10 (21) : http://www.brasilien.de/land/florafauna/tropregenwald.asp	N/A	Ja
11 (78) : http://www.umweltbrief.de/neu/html/archiv/Regenwald.txt	N/A	Ja
12 (18) : http://www.regenwaldschutz.de	5	Ja
13 (17) : http://www.ilhacomprida.net/sitio_das_aguas_cantantes.htm	N/A	Ja
14 (19) : http://www.waldportal.org/books.taiga/print.html	N/A	Ja
15 (22) : http://www.wuestenfahrer.de/westafrika/westafrika_regenwald.htm	N/A	Ja

Tabelle 19: Personalisierte Ergebnisliste des Nutzers VAH der Suchmaschine Fooxx

Bei den Ergebnissen für den Nutzer VAH (Tab. 19) ist neben der Häufung von neun relevanten Seiten unter den ersten zwölf Treffern zu bemerken, dass eine relevante Seite, die sich nicht unter den ersten zehn Treffern befand, nun Platz 1 belegt.

6.1.3 Abfragezeiten der Testreihe I

Die Abfragezeiten für die Abfrage der unpersonalisierten Ergebnisliste von Google und Fooxx lagen zwischen 20sec und 25sec. Die Personalisierungen dieser Ergebnislisten lagen zwischen 1,5sec und 3sec, wobei die Personalisierung eines Treffers, sofern Ähnlichkeitswerte zu diesem Treffer in der Datenbank ermittelt werden konnten, zwischen 150ms und 210ms lag.

6.1.4 Beurteilung von Testreihe I

Insgesamt lässt der Algorithmus eine Verdichtung der von den Testern als relevant bewerteten Seiten unter den ersten Treffern erkennen. Die relevanten Seiten werden teilweise von den ziemlich weit hinten liegenden Plätzen nach vorne gebracht. Es befinden sich noch Seiten unter den Ergebnissen, die von den Testern als nicht so relevant benannt wurden. Trotzdem zeigt der Algorithmus eine überraschend sinnvolle neue Sortierung, da sich nun erheblich mehr relevante Seiten auf den vorderen Plätzen befinden.

Allerdings ist bei Test I zu bemängeln, dass evtl. die Anzahl der Tester und Testreihen zu gering ist. Wünschenswert wäre eine höhere Anzahl an Testern, sowie eine vorherige Erstellung eines möglichst umfassenden Fooxx-Profils der Tester.

Dies könnte im Rahmen eines groß angelegten Tests auf der bestehenden Webseite von Fooxx mit den bestehenden Nutzern erfolgen. Alle bestehenden Nutzer von Fooxx könnten bei Ihrer täglichen Nutzung an diesem Test teilnehmen, indem sie vorhandene Ergebnislisten mit nach dem neuen Algorithmus personalisierten Ergebnislisten vergleichen würden. Dieser Test war im Rahmen dieser Arbeit leider nicht durchführbar.

Ob zudem eine Personalisierung der Treffer vorgenommen wurde, ist in diesem Test nicht direkt erkennbar. Dies ist auf die geringe Anzahl an Testreihen zurückzuführen. Deswegen wird in einem zweiten Schritt die Personalisierung des Algorithmus weitergehend untersucht.

6.2 Testreihe II

Es werden zwei bestehende und bekannte Nutzer von Fooxx genommen, deren Profile bekannt sind. Mit einem vorgegebenen Suchbegriff werden in der Testumgebung die unpersonalisierten Ergebnislisten der Suchmaschinen ermittelt und personalisiert. Die personalisierten Ergebnisse werden mit den Profilen der Nutzer verglichen. Es soll getestet werden, ob wirklich eine Personalisierung durch den Algorithmus nachvollziehbar ist.

6.2.1 Testdurchführung von Testreihe II

In der Testumgebung wird der Suchbegriff „Apple“ eingegeben. (Abb. 10). Die Ergebnisse der Suche liefert fast ausschließlich Seiten des Computer-Herstellers Apple zurückliefert. Dies ist wahrscheinlich am sinnvollsten für den „normalen“ Suchenden, da sich der Infomationswunsch des Nutzers bei solch einem kurzen und nicht eindeutigen Schlüsselwort meist auf diesen Sachverhalt bezieht. Zudem ist dies auch auf die häufigere Verlinkung auf die Homepage des Computerherstellers als auf englischsprachige Seiten zu dem Thema „Apple“ zurückzuführen. Da für diesen Test nicht die Relevanz der Seiten im Vordergrund steht, sondern dessen Personalisierung für den Nutzer, sind die Ergebnisse der Abfrage der für die Testumgebung implementierten Volltextsuche Lucene interessanter. Sie liefert eine bunte Mischung aus Seiten. Auch hier überwiegen Seiten des Computer-Herstellers, allerdings gibt es darüber hinaus ein breites Spektrum an stark unterschiedlichen Seiten.

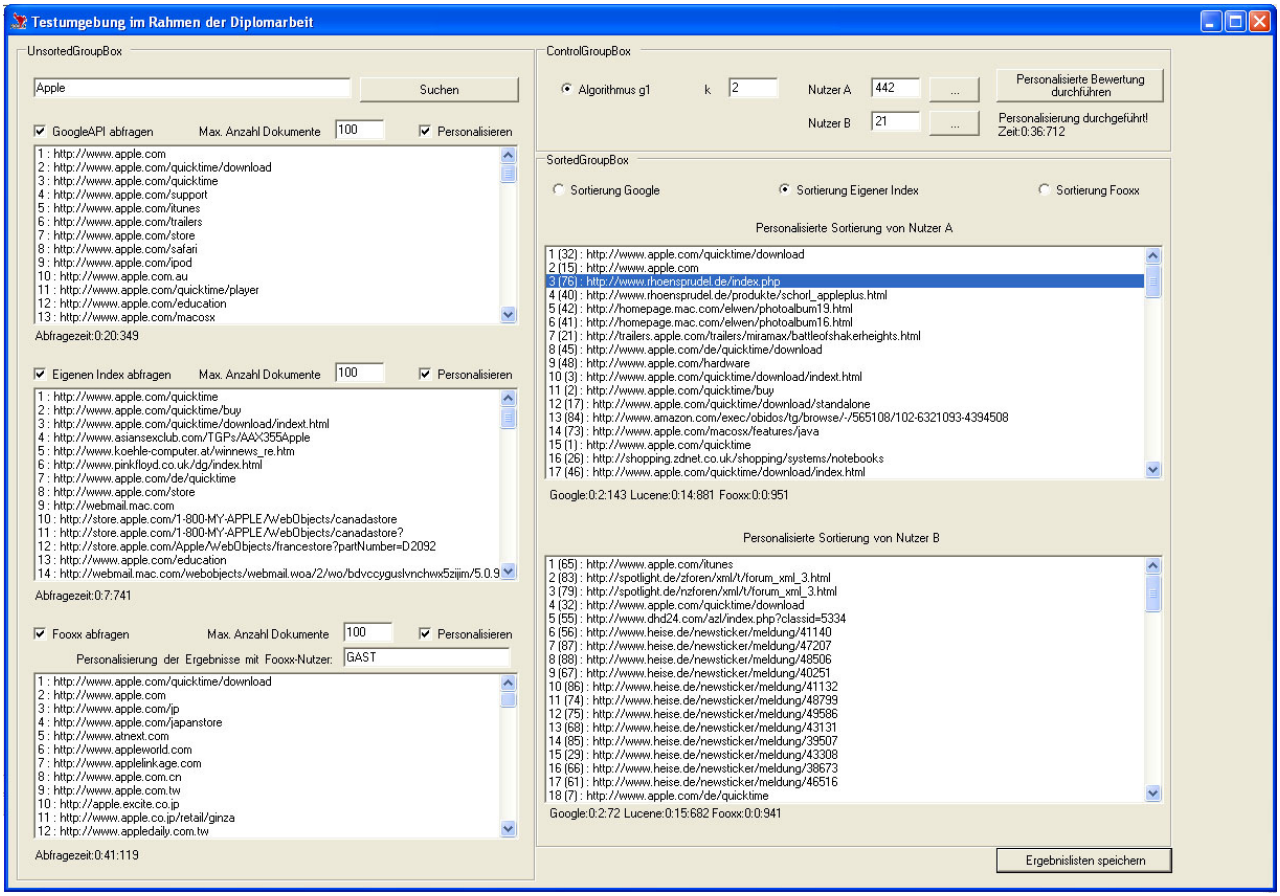


Abbildung 10: Screenshot der Testumgebung bei der Testdurchführung von Testreihe II

Die Personalisierung in der Testumgebung erfolgt mit den Nutzer VAH und PKRUG. Diese beiden sind mit die aktivsten Nutzer von Fooxx. Das Profil dieser beiden ist bekannt, da sie Mitarbeiter der Firma Global Brain Network sind. Die Profile werden wie folgt beschrieben:

- Profil von VAH: Leiter der Entwicklungsabteilung von Fooxx. Sein Schwerpunkt beim Surfen liegt auf technischen Seiten, die sich mit der Thematik Informatik und IT auseinandersetzen.
- Profil von PKRUG: Geschäftsführer von Fooxx. Sein Schwerpunkt beim Surfen liegt auf generellen Informationen als Endanwender sowie Informationen im Bereich Wirtschaft. Zudem interessiert er sich beim privaten Surfen für Seiten mit Informationen über Ernährung.

Die Ergebnislisten vor und nach der Personalisierung sind der Vollständigkeit halber im Anhang B22 aufgeführt.

Da bei der Personalisierung der eigene Index abgefragt wird, kann auch für jedes Dokument ein beschreibender Wert ermittelt werden. Die Ergebnisse der

Personalisierung für die beiden Nutzer finden sich in Tabelle 20 und Tabelle 21, wobei die Werte in Klammern den Rang vor der Personalisierung angeben.

Ergebnisse Personalisierung eigener Index für Nutzer PKRUG und Suchbegriff "Apple"

1 (32) : http://www.apple.com/quicktime/download
2 (15) : http://www.apple.com
3 (76) : http://www.rhoensprudel.de/index.php
4 (40) : http://www.rhoensprudel.de/produkte/schorl_appleplus.html
5 (42) : http://homepage.mac.com/elwen/photoalbum19.html
6 (41) : http://homepage.mac.com/elwen/photoalbum16.html
7 (21) : http://trailers.apple.com/trailers/miramax/battleofshakerheights.html
8 (45) : http://www.apple.com/de/quicktime/download
9 (48) : http://www.apple.com/hardware
10 (3) : http://www.apple.com/quicktime/download/indext.html
11 (2) : http://www.apple.com/quicktime/buy
12 (17) : http://www.apple.com/quicktime/download/standalone
13 (84) : http://www.amazon.com/exec/obidos/tg/browse/-/565108/102-6321093-4394508
14 (73) : http://www.apple.com/macosx/features/java
15 (1) : http://www.apple.com/quicktime
16 (26) : http://shopping.zdnet.co.uk/shopping/systems/notebooks
17 (46) : http://www.apple.com/quicktime/download/index.html
18 (80) : http://spotlight.de/zforen/aphp/m/aphp-1015343642-7165.html
19 (49) : http://www.apple.com/g5processor
20 (82) : http://spotlight.de/zforen/aphp/m/aphp-1015347086-12597.html

Tabelle 20: Ergebnisse der Personalisierung für den Nutzer PKRUG

Ergebnisse Personalisierung eigener Index für Nutzer VAH und Suchbegriff "Apple"

1 (65) : http://www.apple.com/itunes
2 (83) : http://spotlight.de/zforen/xml/t/forum_xml_3.html
3 (79) : http://spotlight.de/nzforen/xml/t/forum_xml_3.html
4 (32) : http://www.apple.com/quicktime/download
5 (55) : http://www.dhd24.com/azl/index.php?classid=5334
6 (56) : http://www.heise.de/newsticker/meldung/41140
7 (87) : http://www.heise.de/newsticker/meldung/47207
8 (88) : http://www.heise.de/newsticker/meldung/48506
9 (67) : http://www.heise.de/newsticker/meldung/40251
10 (86) : http://www.heise.de/newsticker/meldung/41132
11 (74) : http://www.heise.de/newsticker/meldung/48799
12 (75) : http://www.heise.de/newsticker/meldung/49586
13 (68) : http://www.heise.de/newsticker/meldung/43131
14 (85) : http://www.heise.de/newsticker/meldung/39507
15 (29) : http://www.heise.de/newsticker/meldung/43308
16 (66) : http://www.heise.de/newsticker/meldung/38673
17 (61) : http://www.heise.de/newsticker/meldung/46516
18 (7) : http://www.apple.com/de/quicktime
19 (6) : http://www.pinkfloyd.co.uk/dg/index.html
20 (15) : http://www.apple.com

Tabelle 21: Ergebnisse der Personalisierung für den Nutzer PKRUG

Betrachtet man die Ergebnisse nach der Personalisierung (Tab.20 und 21), so wird für die Personalisierung des Nutzers PKRUG (Tab. 20) deutlich, dass sich neben dem bekannten „Quicktime-Plugin“ für ihn als Endanwender und der Homepage des Computer-Herstellers Apple auch auf Rang 3 und 4 zwei Seiten über Apfelsaftschorle

(als Seiten mit Bezug zum Thema Ernährung) befinden. Die restlichen Treffer sind generelle Informationen, die sich für Endanwender eignen.

Für den Nutzer VAH (Tab. 21), dessen Profil den Schwerpunkt IT besitzt, werden in der Ergebnisliste auf den Rängen 2 und 3 sowie 6 bis 17 hauptsächlich Seiten mit Bezug zum Thema IT eingeblendet. Rang 2 und 3 sind Forenbeiträge zum Thema XML, und Rang 6 bis 17 sind Newsmeldungen der bekannten IT-Zeitschrift „Heise“.

Es zeigt sich somit, dass sich je nach Profil des Nutzers eine Verdichtung von für ihn wahrscheinlich relevanten Dokumenten in der Trefferliste bemerkbar macht. Dabei ist zu bemerken, dass die Personalisierung der Dokumente nicht auf die Verwendung durch den Nutzer selbst, sondern auf die Verwendung von anderen Nutzern zurückzuführen ist und somit das in 2.7.5.2 beschriebene Konzept der sozialen Filterung umsetzt.

6.2.2 Auswirkungen von k

Der Faktor *k* wurde eingeführt, um die Personalisierung der Dokumente respektive die Häufigkeit der Aufrufe der Dokumente stärker zu gewichten. Die Auswirkungen von *k* werden im Rahmen dieses Testes genauer untersucht. Dazu werden bei der Personalisierung für *k* die Werte 1, 4 und 10 angenommen, und nur der Nutzer PKRUG weiter betrachtet. Der obige Test wurde bereits mit dem Faktor 2 für den Nutzer PKRUG in Tabelle 16 durchgeführt. Die weiteren Ergebnisse der Personalisierung für den Nutzer PKRUG mit k gleich 1, 4 und 10 finden sich in den Tabellen 22, 23 und 24.

Ergebnisse eigener Index für Nutzer PKRUG, Suchbegriff "Apple" und k gleich 1

1 (15) : http://www.apple.com
2 (45) : http://www.apple.com/de/quicktime/download
3 (32) : http://www.apple.com/quicktime/download
4 (48) : http://www.apple.com/hardware
5 (76) : http://www.rhoensprudel.de/index.php
6 (40) : http://www.rhoensprudel.de/produkte/schorl_appleplus.html
7 (21) : http://trailers.apple.com/trailers/miramax/battleofshakerheights.html
8 (41) : http://homepage.mac.com/elwen/photoalbum16.html
9 (42) : http://homepage.mac.com/elwen/photoalbum19.html
10 (2) : http://www.apple.com/quicktime/buy
11 (1) : http://www.apple.com/quicktime
12 (26) : http://shopping.zdnet.co.uk/shopping/systems/notebooks
13 (17) : http://www.apple.com/quicktime/download/standalone
14 (84) : http://www.amazon.com/exec/obidos/tg/browse/-/565108/102-6321093-4394508
15 (73) : http://www.apple.com/macosx/features/java
16 (3) : http://www.apple.com/quicktime/download/indext.html
17 (46) : http://www.apple.com/quicktime/download/index.html
18 (81) : http://spotlight.de/zforen/aphp/m/aphp-1015346974-12452.html
19 (82) : http://spotlight.de/zforen/aphp/m/aphp-1015347086-12597.html
20 (49) : http://www.apple.com/g5processor

Tabelle 22: Ergebnisse Personalisierung für PKRUG, Suchbegriff „Apple“ und *k* gleich 1

Ergebnisse eigener Index für Nutzer PKRUG, Suchbegriff "Apple" und k gleich 4

1 (32) : http://www.apple.com/quicktime/download
2 (76) : http://www.rhoensprudel.de/index.php
3 (40) : http://www.rhoensprudel.de/produkte/schorl_appleplus.html
4 (42) : http://homepage.mac.com/elwen/photoalbum19.html
5 (21) : http://trailers.apple.com/trailers/miramax/battleofshakerheights.html
6 (41) : http://homepage.mac.com/elwen/photoalbum16.html
7 (15) : http://www.apple.com
8 (45) : http://www.apple.com/de/quicktime/download
9 (17) : http://www.apple.com/quicktime/download/standalone
10 (46) : http://www.apple.com/quicktime/download/index.html
11 (3) : http://www.apple.com/quicktime/download/indext.html
12 (26) : http://shopping.zdnet.co.uk/shopping/systems/notebooks
13 (84) : http://www.amazon.com/exec/obidos/tg/browse/-/565108/102-6321093-4394508
14 (1) : http://www.apple.com/quicktime
15 (73) : http://www.apple.com/macosx/features/java
16 (2) : http://www.apple.com/quicktime/buy
17 (48) : http://www.apple.com/hardware
18 (81) : http://spotlight.de/zforen/aphp/m/aphp-1015346974-12452.html
19 (82) : http://spotlight.de/zforen/aphp/m/aphp-1015347086-12597.html
20 (49) : http://www.apple.com/g5processor

Tabelle 23: Ergebnisse Personalisierung für PKRUG, Suchbegriff „Apple“ und *k* gleich 4

Ergebnisse eigener Index für Nutzer PKRUG, Suchbegriff "Apple" und k gleich 10

1 (1) : http://www.apple.com/quicktime
2 (2) : http://www.apple.com/quicktime/buy
3 (3) : http://www.apple.com/quicktime/download/indext.html
4 (4) : http://www.asiansexclub.com/TGPs/AAX355Apple
5 (5) : http://www.koehle-computer.at/winnews_re.htm
6 (6) : http://www.pinkfloyd.co.uk/dg/index.html
7 (7) : http://www.apple.com/de/quicktime
8 (8) : http://www.apple.com/store
9 (9) : http://webmail.mac.com
10 (10) : http://store.apple.com/1-800-MY-APPLE/WebObjects/canadastore
11 (11) : http://store.apple.com/1-800-MY-APPLE/WebObjects/canadastore?
12 (12) : http://store.apple.com/Apple/WebObjects/francestore?partNumber=D2092
13 (13) : http://www.apple.com/education
14 (14) : http://webmail.mac.com/webobjects/webmail.woa/2/wo/bdvccyguslvnchwx5zijim/5.0.
15 (15) : http://www.apple.com
16 (16) : http://www.apple.com/downloads/macosx/email_chat/smailing.html
17 (17) : http://www.apple.com/quicktime/download/standalone
18 (18) : http://www.apple.com/pr/library/2003/feb/10xserveraid.html
19 (19) : http://www.bestofmicro.com/jhtml2/shop/constructor/constructor_shop-1600042-4900
20 (20) : http://www.apple.com/fr/education

Tabelle 24: Ergebnisse Personalisierung für PKRUG, Suchbegriff „Apple“ und *k* gleich 10

In den Ergebnissen für *k* gleich 10 (Tab. 24) kann man sehen, dass die Personalisierung keine Rolle spielt – die Treffer werden in der Reihenfolge sortiert, wie sie übergeben werden. Dies liegt daran, dass ab einer bestimmten Größe für *k* die Formel als beschreibenden Wert immer 0 zurückliefert. Dies ist mathematisch durch das Bilden des Kehrwerts zu erklären – ab einer bestimmten Größe für *k* werden die beschreibenden Werte so klein, dass sie nicht mehr berechnet werden können.

Beim Faktor *k* gleich 1 (Tab. 22) für die Personalisierung wird im Vergleich mit der Personalisierung mit Faktor *k* gleich 2 deutlich, dass die Treffer zum Thema Ernährung beim Nutzer PKRUG, die eher zu seinem Profil passen, von Platz 3 und 4 auf Platz 5 und 6 „abrutschen“. Wählt man dem Faktor *k* gleich 4 (Tab. 23), so landen die beiden Seiten auf den Plätzen 2 und 3.

Somit wird deutlich, dass bei einem Faktor für *k* gleich 1 die Personalisierung und die Häufigkeit des Vorkommens gleich gewichtet sind – je größer *k* gewählt wird, umso mehr tritt die Personalisierung in den Vordergrund. Ab einem bestimmten Wert für *k* ist eine Berechnung des beschreibenden Wertes nicht mehr möglich.

Mathematisch könnte gezeigt werden, dass mit Werten für *k*, die im Bereich 0 bis 1 liegen, gleiche Ergebnisse erreicht werden. Im Rahmen dieser Ausarbeitung wird dies nicht weiter verfolgt.

6.2.3 Abfragezeiten

Da der eigene Index lokal gespeichert ist, lag dessen Abfrage bei ca. 7sec. Die Personalisierung eines Treffers lag zwischen 150ms und 220ms.

6.3 Abschließende Beurteilung von Testreihe I und Testreihe II

In Testreihe II sieht man, dass auch eine Personalisierung vorgenommen wird. Dies kann durch den Faktor k entsprechend beeinflusst werden. Es ist anzunehmen, dass der vorgestellte Personalisierungsalgorithmus seinen Aufgaben entspricht, die gewünschten Resultate liefert und die Treffer nach dem Profil des suchenden Nutzers sortiert.

Nachdem nur der Algorithmus auf seine Funktion hin untersucht wurde, wird nun versucht, eine kürzere Berechnungszeit für die Personalisierung eines Dokumentes zu erreichen, da die momentane Berechnungszeit noch recht hoch ist.

7 Durchgeführte Optimierungen

Die Abfragedauer einer Suchmaschine ist einer der kritischsten Punkte. Meist müssen innerhalb weniger Sekunden bzw. Millisekunden aus einer sehr großen Anzahl an Dokumenten die relevanten Dokumente gefunden werden. Diese müssen dann möglichst schnell nach deren Relevanz sortiert und wiedergegeben werden. Deswegen sollte eine möglichst kurze Dauer des Findens und Sortierens der Dokumente angestrebt werden.

7.1 Optimierungsansätze

Der hier vorgestellte Personalisierungsalgorithmus benötigt zwischen 150ms und 220ms für das Ermitteln eines beschreibenden Wertes eines Dokumentes. Sobald alle beschreibenden Werte ermittelt sind, können die Dokumente dann mit Hilfe eines geeigneten Sortierungsalgorithmus (z.B. *QuickSort*) sortiert werden. Solche Sortierungsalgorithmen sind in der Literatur beschrieben, sehr schnell und effizient und werden hier nicht näher erläutert, da sie keine weitere Optimierung benötigen.

In der Testumgebung wurden Zeitmessungen durchgeführt, welche die Abfrage der entsprechenden Daten aus der Datenbank sowie die Dauer der Berechnung der Relevanz eines Dokumentes mit dem entwickelten Algorithmus festhielten. Bei der Betrachtung der Messungen (z.B. Abb. 8 in Kapitel 5.8.3) fällt auf, dass ausschließlich die Datenbankabfrage eine Rolle spielte. Dabei stellt sich heraus, dass die Berechnung des beschreibenden Wertes innerhalb einer Prozessorzeitscheibe, die dem Programm zugeteilt wurde, durchgeführt wurde. Die Zeitdauer der Berechnung liegt somit unterhalb des Millisekundenbereichs. Die restliche Zeitdauer der Personalisierung nimmt die Datenbankabfrage in Anspruch.

Die Datenbank-Abfrage wurde bereits in 5.10.1 sowie in Quellcode 2 und 3 dargestellt und näher erläutert. Es wird nun vorgestellt, welche Optimierungen vorgenommen wurden, um diese Abfrage zu verbessern.

7.2 Optimierung der Datenbankstruktur- und abfrage

Zunächst wurde die Datenbankstruktur verbessert. Normalerweise müssten die Ähnlichkeitswerte eines jeden Nutzers „just in time“ in der Datenbank ausgelesen werden. Ein Job wurde eingerichtet, der diese Ähnlichkeitswerte in Intervallen

nacheinander für alle Nutzer im Voraus berechnet und direkt in eine eigene Tabelle schreibt. Dieser Job aktualisiert somit ständig die Ähnlichkeitstabelle. (Die so vorbereitete Tabelle heißt `similarity` und findet sich im in Kapitel 5.10.1 in Abbildung 9) Dieser Job läuft momentan auf dem Fooxx-Server und braucht für eine komplette Aktualisierung aller Ähnlichkeitswerte ungefähr 1 Woche (bei über 160.000 Dokumenten und 800 registrierten Fooxx-Nutzern). Dies hat zwar den Nachteil, dass die abgefragten Ähnlichkeitswerte nicht aktuell sind, dafür erniedrigt sich die Abfragezeit. Die Optimierung wurde bereits vor der Entwicklung der Testumgebung implementiert. Die in Kapitel 6 angegebenen Abfragezeiten sind somit schon teilweise optimierte Zeiten.

Weiterhin wurde versucht, die Dauer der Abfrage an die Datenbank zu senken. Bei der in 5.10.1 und in Quellcode 2 und 3 vorgestellten Lösung wird für jedes Dokument eine Abfrage an die Datenbank geschickt und die Antwort berechnet. Bei der Optimierung bei Datenbankabfragen lohnt es sich in der Regel, meist eine möglichst umfangreiche Abfrage an die Datenbank zu schicken, da dann die Datenbank selbst Optimierungen bei dieser Abfrage vornehmen kann. So wurde versucht, die Abfragedauer dadurch zu optimieren, dass die Abfrage der Ähnlichkeitswerte von allen zu personalisierenden Dokumenten in einer einzelnen Abfrage übergeben wurde. Dadurch wurde die in Quellcode 3 dargestellte Abfrage mit dem *SQL*-Befehl „UNION“ zu einer neuen „großen“ Abfrage zusammengefasst. Anschließend wurden die Ähnlichkeitswerte der einzelnen Dokumente aus der zurück gelieferten Tabelle wieder ausgelesen und berechnet. Allerdings stellte sich leider der gegenteilige Effekt heraus: Vorher dauerte eine Personalisierung von 100 Dokumenten ca. 15sec. Nach der Optimierung dauerte die Personalisierung 20sec. Die Optimierung war somit nicht erfolgreich.

Die Berechnungszeit für eine Abfrage liegt somit nach allen durchgeführten und versuchten Optimierungen weiterhin bei 150ms bis 220ms. Es wäre wünschenswert, weitere Optimierungen der Datenbankabfrage sowie der Datenbankstruktur durchzuführen, was im Rahmen dieser Arbeit jedoch nicht mehr möglich war.

8 Ergebnis der Arbeit

Im Rahmen dieser Arbeit wurden zunächst die Grundlagen des Information Retrieval dargelegt und insbesondere benutzerverhaltensorientierte Ansätze dargestellt. Die bestehende benutzerverhaltensorientierte Suchmaschine Fooxx und deren Personalisierungsmethode wurde beschrieben. Anschließend ist ein neuer Relevanzbewertungs-Algorithmus entwickelt worden, den es in dieser Form noch nicht gab. Er unterliegt dem Prinzip der sozialen Filterung und berücksichtigt die Beziehungen zwischen verschiedenen Nutzern und deren Verhalten. Weiterhin wurde eine Testumgebung entwickelt, mit der zunächst der entwickelte Algorithmus getestet werden konnte. Zudem könnten weitere Algorithmen mit dieser Testumgebung untersucht werden. Es wurden Tests durchgeführt, die den Algorithmus auf seine Funktionsweise hin überprüften. Zum Schluss wurde noch erläutert, welche Optimierungen beim Algorithmus gemacht wurden und ggf. anzustreben sind.

Die Testumgebung selbst bildet dabei ein Auswertungstool, das mit vielen Funktionen einen guten Einblick in den bestehenden Algorithmus bietet. Durch die Anbindung an die Suchmaschinen Fooxx und Google, sowie an den für diese Arbeit erstellten eigenen Index von Fooxx können verschiedene Dokumentenmengen mit dem entwickelten und auch mit weiteren Algorithmen getestet werden.

In den Testreihen wurde der Algorithmus mit Hilfe der Testumgebung auf seine Aufgabenerfüllung hin untersucht. Es wurde festgestellt, das der Algorithmus die relevanten Seiten einer großen Dokumentenmenge (100 Dokumente), die in Ihren Rängen etwa gleich verteilt sind, entsprechend als relevanter bewertet und bei der anschließenden Sortierung diese relevanten Dokumente auf die vorderen Ränge bringt. Weiterhin wurde in einer zweiten Testreihe gezeigt, dass der Algorithmus eine Personalisierung der Ergebnisse für den jeweiligen Nutzer vornimmt. Mit dem im Algorithmus angegebenen Parameter *k* kann die Personalisierung des Algorithmus wie gewünscht beeinflusst werden.

Zum Schluss wurden die vorgenommenen Optimierungen beschrieben. Da ausschließlich die Datenbankabfrage optimiert werden kann, wurden zwei Optimierungsschritte vorgenommen. Der erste wurde bereits bei der bestehenden

Testumgebung und Fooxx-Datenbank implementiert, der zweite brachte nicht den gewünschten Erfolg. Weitere Optimierungen der Datenbank und der Datenbankstruktur sind wünschenswert, da die Berechnungszeit für ein zu personalisierendes Dokument mit 150ms bis 220ms noch relativ hoch ist.

Der Algorithmus scheint seinen Zweck zu erfüllen. Wie schon in der Einleitung und in Kapitel 3 erwähnt, ist es geplant, ihn in das bestehende Fooxx als Ersatz für die bestehende Personalisierung und Profilerstellung zu integrieren. Dies sollte im Rahmen eines größer angelegten Tests folgendermaßen vorgenommen werden: Auf einer Testseite sollte der Algorithmus die bestehende Profilerstellung aufgrund der Gemeinschaften sowie die anschließende Personalisierung ersetzen und zusammen mit der bestehenden Grundbewertung der Seiten und der bestehenden Sortierung anhand des „inneren“ Werts implementiert werden. Ausgewählte Fooxx-Nutzer bekommen dann Zugang zu diesen Test-Seiten, können auf diesen die „neue“ Suche mit der „alten“ Suche vergleichen und dazu Ihre Bewertung und Meinung abgeben.

Auf diese Weise wäre es möglich, weitere Aussagen über die Qualität des Algorithmus zu erhalten. Zudem könnten im Rahmen dieses Tests weitere Optimierungen in Bezug auf die Abfragezeit der Datenbank vorgenommen werden. Aufgrund der bisherigen Testreihen ist zu erwarten, dass im Rahmen des weitergehenden Tests der Algorithmus zeigen wird, dass er die gewünschten Anforderungen erfüllt und somit die bestehende Profilerstellung und Personalisierung in Fooxx ersetzen kann.

Über das Konzept des entwickelten Algorithmus selbst lässt sich festhalten, dass die implizite Form der Profilerstellung einen großen Vorteil darstellt. Viele Algorithmen, die auf dem Konzept der sozialen Filterung bestehen, benutzen eine manuelle Profilerstellung oder eine halb-manuelle Profilerstellung mit Computerunterstützung. Der im Rahmen dieser Arbeit entwickelte Personalisierungs-Algorithmus zeichnet sich durch seine komplett automatisierte Profilerstellung aus, die durch ein entsprechendes Tool (z.B. der Fooxx-Toolbar) umgesetzt wird. Denkbar sind auch andere Tools, mit denen diese automatisierte Profilerstellung unterstützt werden kann. So könnte z.B. ein Betreiber eines Forum-Systems im Internet eine

automatische Profilerstellung seiner Nutzer durchführen, indem er dafür die Log-Dateien über die aufgerufenen Forum-Beiträge der Nutzer als Basis nimmt.

Somit sind weitergehende Einsatzbereiche für diesen Algorithmus denkbar – so kann damit jedes beliebige Suchsystem erweitert werden, welches in irgendeiner Form Dokumente, Nutzer und deren Beziehungen zueinander vorhält. Denkbar wäre z.B. eine Erweiterung der Suche in dem Forum-System: Dort könnte dann ein Nutzer gezielt nach Beiträgen suchen - die Ergebnisliste würde ihm Beiträge zuerst anzeigen, welche von ähnlichen Nutzern bereits verwendet wurden. Gerade für Foren, in denen unterschiedliche Themenbereiche behandelt werden, wäre diese Personalisierung vorteilhaft. Der Suchende würde Ergebnisse in dem Themenbereich zuerst angezeigt bekommen, in dem ähnliche Nutzer bereits häufig Beiträge aufgerufen haben. Dies ist dann der Themenbereich, der seinem Profil am meisten entspricht und den Suchenden wahrscheinlich auch am meisten interessiert.

Abschließend kann gesagt werden, dass in dieser Arbeit erfolgreich in Konzept zur Personalisierung einer Relevanzbewertung für Dokumente entwickelt, implementiert und getestet wurde. Diese Relevanzbewertung kann z.B. in den bestehenden Suchdienst Fooxx integriert werden, aber auch andere Einsatzbereiche sind denkbar.

Quellenverzeichnis

[1] Fang Liu, Clement Yu, Weiyi Meng: *Personalized Web Search for Improving Retrieval Effectiveness*. Department of Computer Science, University of Illinois at Chicago, Internationale Patentanmeldung

[2] Martin Heß: *Verteiltes Information Retrieval für nicht-kooperative Suchserver im WWW.* Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften, vorgelegt beim Fachbereich Biologie und Informatik der Johann Wolfgang Goethe-Universität in Frankfurt/Main, 2002.

[3] Peter Krug: *Verfahren und Vorrichtung zur Ermittlung von relevanten Objekten*. Internationale Patentanmeldung, DE-Patentanmeldung Nr. 101 43 940.7 vom 7.9.2001

[4] Matt Loney: Googles Technologien: *Von Zauberei kaum zu unterscheiden.* ZDNet, 31. Januar 2005.
http://www.zdnet.de/itmanager/unternehmen/0,39023441,39129811-2,00.htm
Quelle auf beiliegender CD: „Quellen/Q4/Q4_seite1.htm“

[5] Annabel Pollock, Andrew Hockley: *What's wrong with Internet searching*.
D-Lib Magazine, March 1997, ISSN 1082-9873
http://www.dlib.org/dlib/march97/bt/03pollock.html
Quelle auf beiliegender CD: „Quellen/Q5/Q5.htm“

[6] N.J.Davies, M.C.Revett: *Networked information management*
Internationale Patentanmeldung XP-00703570, BT Technol. J Vol. 15 No.2, April 1997

[7] Verschieden Autoren: *Definition Information Retrieval in Wikipedia*
http://de.wikipedia.org/wiki/Information-Retrieval
Quelle auf beiliegender CD: „Quellen/Q7/Q7.htm“

[8] Lancaster,F.W: *Information Retrieval Systems: Characteristics, Testing and Evaluation.* Wiley, New York (1968)

[9] C.J. van Rijsbergen: *Information Retrieval* (2nd Edition). Butterworths, 1975

[10] Dominik Kuropka: Modell Information Retrieval.
http://de.wikipedia.org/wiki/Bild:Information-Retrieval.ng
Quelle auf beiliegender CD: „Quellen/Q10/Q10.htm“

[11] H. Winkler. *Suchmaschinen – Metamedien im Internet? Telepolis*, 1997.
http://wwwcs.uni-paderborn.de/~winkler/suchm_d.html
Quelle auf beiliegender CD: „Quellen/Q11/Q11.htm“

[12] Bernard Bekavac: *Methoden und Verfahren von Suchdiensten im WWW/Internet*
http://www.inf-wiss.uni-konstanz.de/suche/tutorial/such_tutorial_advanced.html
Quelle auf beiliegender CD: „Quellen/Q12/Q12.htm“

[13] Brian H. Murray, Cyveillance Inc.: *Sizing the Internet - A Cyveillance White Paper.* July 2000.
http://www.cyveillance.com/web/downloads/Sizing_the_Internet.pdf
Quelle auf beiliegender CD: “ Quellen/Q13/Q13.pdf”

[14] Greg R. Notess: *Search Engine Statistics – Database Total Size Estimates*
http://www.notess.com/search/stats/sizeest.shtml
Quelle auf beiliegender CD: „Quellen/Q14/Q14.htm“

[15] Michael K. Bergmann: *The Deep Web – Surfacing Hidden Value.*
Deep Content White Paper
http://www.brightplanet.com/pdf/deepwebwhitepaper.pdf
Quelle auf beiliegender CD: “ Quellen/Q15/Q15.pdf”

[16] Sebastian Wolf: *Statistik: Aktuelle Zahlen zum WWW.*
Universitätsbibliothek Bielefeld, 4.1.2005
http://www.ub.uni-bielefeld.de/biblio/search/help/statistik.htm
Quelle auf beiliegender CD: „Quellen/Q16/Q16.htm“

[17] Iiya Geller: *Generating personalized user profiles for utilizing the generated user profiles to perform adaptive internet searches*. Eingereicht durch Mightiest Logicon Unisearch inc., Internationale Patentanmeldung, Patentnr.: PCT/US00/01373 ,. 20.1.2000

[18] G. Salton: *The SMART Retrieval System – Experiments in Automatic Document Processing*. Prentice Hall Inc., 1971.

[19] Projekt deutscher Wortschatz: Häufigkeitsklassen.
http://wortschatz.uni-leipzig.de/html/faq/hkl.html
Quelle auf beiliegende CD: „Quellen/Q19/Q19_1.htm“
Damián H. Zanette: *Zipf's law and the creation of musical context.* August 2004.
http://arxiv.org/PS_cache/cs/pdf/0406/0406015.pdf
Quelle auf beiliegender CD: „Quellen/Q19/Q19_2.pdf“

[20] Danny Sullivan: *Google Loses Tabs In New Look, Gains Web Alerts & Personalized Search Results*
http://searchenginewatch.com/searchday/article.php/3332511
Quelle auf beiliegender CD: „Quellen/Q20/Q20.htm“

[21] Michael Weiss: *Verfahren zum Durchsuchen elektronisch gespeicherter Dokumente.* Mitel Corp, Kanada, Ontario, Deutsche Patenrechtsanmeldung DE 199 13 509 A1

[22] Yezdi Lashkari, Max Metral, Pattie Maes: *Collaborative Interface Agents.*
http://agents.www.media.mit.edu/groups/agents/publications/aaai-ymp/aaai.html
Quelle auf beiliegender CD “ Quellen/Q22/Q22_seite1.htm”

Anhang

A. Quellcode

A1: Code der Methode `UseGoogleAPI` zur Abfrage der Ergebnisse von Google Befindet sich auf beiliegender CD unter „TestEnvironment/TestAPI.cs“, Zeilen 89-138

A2: Code der Methode `UseFooxxAPI` zur Abfrage der Ergebnisse von Fooxx. Befindet sich auf beiliegender CD unter „TestEnvironment/TestAPI.cs“, Zeilen 140-212

A3: Code der Methode `UseLuceneIndex` zur Abfrage des eigenen Index. Befindet sich auf beiliegender CD unter „TestEnvironment/TestAPI.cs“, Zeilen 214-273

A4: Kompletter Quellcode der Packages `TestEvironment` befindet sich auf der beiliegenden CD im Ordner „TestEnvironment“. Der komplette Quellcode des Packages `RankAlgorithm` befindet sich auf der beiliegenden CD im Ordner „RankAlgorithm“.

A5: Die Ausführbaren Dateien befinden sich auf der beiliegenden CD im Ordner „Bin“.

B. Aufgabenblatt und Ergebnisse des Tests

B1: Aufgabenblatt, welches von den Testern zu bearbeiten war

Aufgabenblatt

Bitte bearbeite die folgenden Aufgaben der Reihe nach.

1) Finde Informationen über die gesundheitlichen Vorzüge von Tee. Benutze dazu die Schlüsselworte „Vorzüge von Tee“ (ohne Anführungszeichen) in der Suchmaschine Fooxx. Die URL von Fooxx lautet: www.fooxx.com. Gib die 5 Seiten an, die deiner Meinung nach das Thema am besten beschrieben haben und füge sie den Favoriten hinzu.

2) Finde Informationen über die Auswirkungen der Abholzung des Regenwalds. Benutze dazu die Schlüsselworte „Abholzung Regenwald“ in der Suchmaschine Google. Die URL von Google lautet: www.google.de. Gib die 5 Seiten an, die deiner Meinung nach das Thema am besten beschrieben haben, und füge Sie den Favoriten hinzu.

Vielen Dank für die Teilnahme.

B2: Testergebnisse der Aufgabe 1 des Aufgabenblatts

Tester	Relevante Seiten der Ergebnisse von "Vorzüge von Tee" bei Fooxx	Hfgk.
Tester 1	http://www.tee-special.com/tee.html	9
Account:	http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	6
ROADRUNNERLENNY	http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	3
	http://teeverband.at/wiss-sort01.htm	9
	http://www.wendtour.de/ernaehrungtee.htm	1
Tester 2	http://www.tee-special.com/tee.html	9
Account:	http://www.edeka.de/EDEKA/Content/DE/ForYou/EssenWohlfuehlen/Ernaehrungst	2
ROADRUNNERLENNY	http://www.flexonline-de.com/183.html	3
	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	7
	http://teeverband.at/wiss-sort01.htm	9
Tester 3	http://www.tee-special.com/tee.html	9
Account:	http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	6
ROADRUNNERLENNY	http://www.edeka.de/EDEKA/Content/DE/ForYou/EssenWohlfuehlen/Ernaehrungst	2
	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	7
	http://www.wander.ch/cgi/de/magazine/contact_archive/contact_21/contact1.asp	1
Tester 4	http://deutschesfachbuch.de/info/detail.php?isbn=3453141784	2
Account:	http://www.neuesleben.net/archive/text/gruentee01.shtml	2
VAH	http://www.frauke-spehr.info/	3
	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	7
	http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	5
Tester 5	http://teeverband.at/wiss-sort01.htm	9
Account:	http://www.tee-special.com/tee.html	9
VAH	http://www.teebuch.de/teebuch.txt	4
	http://www.pandorahs-traum.de/schatztruhe/tee.html	2
	http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	5
Tester 6	http://www.abnehmtreff.de/article133.html	2
Account:	http://www.tee-special.com/tee.html	9
DASDINGSDA	http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	6
	http://www.dreisesselapotheke.de/htm/ernaehrung/99_10-61_7548_TeeZeit_2_200	2
	http://teeverband.at/wiss-sort01.htm	9
Tester 7	http://www.abnehmtreff.de/article133.html	2
Account:	http://de.isodisnatura.ch/nutrition_-_article.htm?ID=14	2
DASDINGSDA	http://www.flexonline-de.com/183.html	3
	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	7
	http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	3
Tester 8	http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	6
Account:	http://www.flexonline-de.com/183.html	3
DASDINGSDA	http://www.teebuch.de/teebuch.txt	4
	http://www.swr.de/kaffee-oder-tee/essen/tee/2003/04/16/	3
	http://teeverband.at/wiss-sort01.htm	9
Tester 9	http://www.tee-special.com/tee.html	9
Account:	http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	6
JKESSLER	http://www.swr.de/kaffee-oder-tee/essen/tee/2003/04/16/	3
	http://www.teebuch.de/teebuch.txt	4
	http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	5

Tester	Relevante Seiten der Ergebnisse von "Vorzüge von Tee" bei Fooxx	Hfgk.
Tester 10	http://www.tee-special.com/tee.html	9
Account:	http://www.frauke-spehr.info/	3
JKESSLER	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	7
	http://teeverband.at/wiss-sort01.htm	9
	http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	5
Tester 11	http://www.teebuch.de/005008.htm	2
Account:	http://teeverband.at/wiss-sort01.htm	9
JKESSLER	http://www.pralinen-teeke.de/shop/html/tee.html	1
	http://www.frauke-spehr.info/	3
	http://www.pandorahs-traum.de/schatztruhe/tee.html	2
Tester 12	http://www.tee-special.com/tee.html	9
Account:	http://www.teeimport.com/teeabc.htm	1
VAH	http://www.teebuch.de/teebuch.txt	4
	http://www.geocities.com/Athens/Delphi/3967/tee.html	1
	http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	3
Tester 13	http://www.teebuch.de/005008.htm	2
Account:	http://www.tee-special.com/tee.html	9
DASDINGSDA	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	7
	http://de.isodisnatura.ch/nutrition_-_article.htm?ID=14	2
	http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	5
Tester 14	http://www.swr.de/kaffee-oder-tee/essen/tee/2003/04/16/	3
Account:	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	7
ROADRUNNERLENNY	http://www.steffys.de/htm/gruenertee.htm	1
	http://teeverband.at/wiss-sort01.htm	9
	http://www.dreisesselapotheke.de/htm/ernaehrung/99_10-61_7548_TeeZeit_2_200	2
Tester 15	http://teeverband.at/wiss-sort01.htm	9
Account:	http://www.neuesleben.net/archive/text/gruentee01.shtml	2
DASDINGSDA	http://www.kapstadt-news.de/news/283.htm	1
	http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	6
	http://deutschesfachbuch.de/info/detail.php?isbn=3453141784	2

B3: Testergebnisse der Aufgabe 1 des Aufgabenblatts

Tester	Relevante Seiten der Ergebnisse von "Abholzung Regenwald" bei Google	Hfgk.
Tester 1	http://www.abenteuer-regenwald.de/abholz.php	2
Account:	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	9
ROADRUNNERLENNY	http://www.regenwald.org/new/regenwaldreport/artikel.php?id=147	2
	http://www.regenwald.org/new/amazonas/highnoon.htm	9
	http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	3
Tester 2	http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	5
Account:	http://www.der-gruene-faden.de/text/text648.html	5
ROADRUNNERLENNY	http://www.faszination-regenwald.de/	2
	http://www.regenwald-spende.de/ueber_uns.htm	5
	http://www.regenwald.org/new/amazonas/highnoon.htm	9
Tester 3	http://berg.heim.at/tibet/450508/Regen.htm	3
Account:	http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms/	4
ROADRUNNERLENNY	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	9
	http://www.regenwald-institut.de/deutsch/Aktionen/Bolivien.htm	2
	http://www.umg.at/112001/abholzung.php	3
Tester 4	http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	3
Account:	http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	5
VAH	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	9
	http://www.regenwald.org/new/amazonas/highnoon.htm	9
	http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	5
Tester 5	http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms/	4
Account:	http://www.der-gruene-faden.de/text/text648.html	5
VAH	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	9
	http://www.regenwald-spende.de/ueber_uns.htm	5
	http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	5
Tester 6	http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	3
Account:	http://www.regenwald.org/pdf/rdr-report0401.pdf#search='Abholzung%20Regenwald'	1
DASDINGSDA	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	9
	http://www.umweltlexikon-online.de/fp/archiv/RUBnaturartenschutz/Regenwald.php	2
	http://www.regenwaldschutz.de/	2
Tester 7	http://www.krref.krefeld.schulen.net/referate/erdkunde/r0235t00.htm	1
Account:	http://www.hauptschule-gochsheim.de/projekte/reg_auswirk.php	1
DASDINGSDA	http://www.regenwald.org/new/amazonas/highnoon.htm	9
	http://www.umweltlexikon-online.de/fp/archiv/RUBnaturartenschutz/Regenwald.php	2
	http://www.cil-frankfurt.de/programme/NEINN/MixMax/Dokumente/CRRegenwald.htm	2
Tester 8	http://www.der-gruene-faden.de/text/text648.html	5
Account:	http://www.econautix.de/site/econautixpage_46.php	4
DASDINGSDA	http://www.regenwald.org/new/amazonas/highnoon.htm	9
	http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	3
	http://www.cil-frankfurt.de/programme/NEINN/MixMax/Dokumente/CRRegenwald.htm	2
Tester 9	http://www.econautix.de/site/econautixpage_46.php	4
Account:	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	9
JKESSLER	http://www.regenwald-institut.de/deutsch/Aktionen/Bolivien.htm	2
	http://www.regenwald.org/new/amazonas/highnoon.htm	9
	http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	5

Tester	Relevante Seiten der Ergebnisse von "Abholzung Regenwald" bei Google	Hfgk.
Tester 10	http://berg.heim.at/tibet/450508/Regen.htm	3
Account:	http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	5
JKESSLER	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	9
	http://www.regenwald.org/new/amazonas/highnoon.htm	9
	http://www.umg.at/112001/abholzung.php	3
Tester 11	http://www.regenwald-spende.de/ueber_uns.htm	5
Account:	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	9
JKESSLER	http://www.econautix.de/site/econautixpage_46.php	4
	http://www.faszination-regenwald.de/	2
	http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms/	4
Tester 12	http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms/	4
Account:	http://www.regenwald.org/new/amazonas/highnoon.htm	9
VAH	http://www.regenwald-spende.de/ueber_uns.htm	5
	http://www.econautix.de/site/econautixpage_46.php	4
	http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	5
Tester 13	http://www.regenwald-spende.de/ueber_uns.htm	5
Account:	http://www.regenwald.org/new/amazonas/highnoon.htm	9
DASDINGSDA	http://www.der-gruene-faden.de/text/text648.html	5
	http://www.umg.at/112001/abholzung.php	3
	http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	5
Tester 14	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	9
Account:	http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	5
ROADRUNNERLENNY	http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	3
	http://www.regenwaldschutz.de/	2
	http://berg.heim.at/tibet/450508/Regen.htm	3
Tester 15	http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	5
Account:	http://www.regenwald.org/new/regenwaldreport/artikel.php?id=147	2
DASDINGSDA	http://www.der-gruene-faden.de/text/text648.html	5
	http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	3
	http://www.abenteuer-regenwald.de/abholz.php	2

B4: Rangfolge der relevanten Seiten von Aufgabe 1 (ohne Unterscheidung der Fooxx-Nutzer)

Platz	Rangfolge der relevanten Seiten (für alle Nutzer) für "Vorzüge von Tee"	Hfgk.
1	http://www.tee-special.com/tee.html	9
1	http://teeverband.at/wiss-sort01.htm	9
2	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	7
3	http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	6
4	http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	5
5	http://www.teebuch.de/teebuch.txt	4
6	http://www.flexonline-de.com/183.html	3
6	http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	3
6	http://www.frauke-spehr.info/	3
7	http://www.swr.de/kaffee-oder-tee/essen/tee/2003/04/16/	3
7	http://www.edeka.de/EDEKA/Content/DE/ForYou/EssenWohlfuehlen/Ernaehrungstipps/Trinkg	2
7	http://www.abnehmtreff.de/article133.html	2
7	http://www.teebuch.de/005008.htm	2
7	http://deutschesfachbuch.de/info/detail.php?isbn=3453141784	2
7	http://www.neuesleben.net/archive/text/gruentee01.shtml	2
7	http://www.pandorahs-traum.de/schatztruhe/tee.html	2
7	http://www.dreisesselapotheke.de/htm/ernaehrung/99_10-61_7548_TeeZeit_2_2003.pdf	2
7	http://de.isodisnatura.ch/nutrition_-_article.htm?ID=14	2
8	http://www.pralinen-teeke.de/shop/html/tee.html	1
8	http://www.wander.ch/cgi/de/magazine/contact_archive/contact_21/contact1.asp	1
8	http://www.wendtour.de/ernaehrungtee.htm	1
8	http://www.teeimport.com/teeabc.htm	1
8	http://www.steffys.de/htm/gruenertee.htm	1
8	http://www.geocities.com/Athens/Delphi/3967/tee.html	1
8	http://www.kapstadt-news.de/news/283.htm	1

B5: Rangfolge der Häufigkeit von Aufgabe 1 für den Nutzer ROADRUNNERLENNY

Platz	Rangfolge der relevanten Seiten von Aufgabe 1 für ROADRUNNERLENNY	Hfgk.
1	http://www.tee-special.com/tee.html	3
1	http://teeverband.at/wiss-sort01.htm	3
1	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	3
2	http://www.edeka.de/EDEKA/Content/DE/ForYou/EssenWohlfuehlen/Ernaehrungstipps/Trinkgenuss	2
2	http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	2
3	http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	1
3	http://www.wendtour.de/ernaehrungtee.htm	1
3	http://www.flexonline-de.com/183.html	1
3	http://www.wander.ch/cgi/de/magazine/contact_archive/contact_21/contact1.asp	1
3	http://www.swr.de/kaffee-oder-tee/essen/tee/2003/04/16/	1
3	http://www.steffys.de/htm/gruenertee.htm	1
3	http://www.dreisesselapotheke.de/htm/ernaehrung/99_10-61_7548_TeeZeit_2_2003.pdf	1

B6: Rangfolge der Häufigkeiten von Aufgabe 1 für den Nutzer VAH

Platz	Rangfolge der relevanten Seiten von Aufgabe 1 für VAH	Hfgk.
1	http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	2
1	http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	2
1	http://www.teebuch.de/teebuch.txt	2
1	http://www.tee-special.com/tee.html	2
2	http://deutschesfachbuch.de/info/detail.php?isbn=3453141784	1
2	http://www.neuesleben.net/archive/text/gruentee01.shtml	1
2	http://www.frauke-spehr.info/	1
2	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	1
2	http://teeverband.at/wiss-sort01.htm	1
2	http://www.pandorahs-traum.de/schatztruhe/tee.html	1
2	http://www.teeimport.com/teeabc.htm	1
2	http://www.geocities.com/Athens/Delphi/3967/tee.html	1
2	http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	1

B7: Rangfolge der Häufigkeiten von Aufgabe 1 für den Nutzer DASDINGSDA

Platz	Rangfolge der relevanten Seiten von Aufgabe 1 für Nutzer DASDINGSDA	Hfgk.
1	http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	3
1	http://teeverband.at/wiss-sort01.htm	3
2	http://www.abnehmtreff.de/article133.html	2
2	http://www.flexonline-de.com/183.html	2
2	http://www.tee-special.com/tee.html	2
2	http://de.isodisnatura.ch/nutrition_-_article.htm?ID=14	2
2	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	2
3	http://www.dreisesselapotheke.de/htm/ernaehrung/99_10-61_7548_TeeZeit_2_2003.pdf	1
3	http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	1
3	http://www.teebuch.de/teebuch.txt	1
3	http://www.swr.de/kaffee-oder-tee/essen/tee/2003/04/16/	1
3	http://www.neuesleben.net/archive/text/gruentee01.shtml	1
3	http://www.kapstadt-news.de/news/283.htm	1
3	http://deutschesfachbuch.de/info/detail.php?isbn=3453141784	1
3	http://www.teebuch.de/005008.htm	1
3	http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	1

B8: Rangfolge der Häufigkeiten von Aufgabe 1 für den Nutzer JKESSLER

Platz	Rangfolge der relevanten Seiten von Aufgabe 1 für Nutzer JKESSLER	Hfgk.
1	http://www.tee-special.com/tee.html	2
1	http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	2
1	http://www.frauke-spehr.info/	2
1	http://teeverband.at/wiss-sort01.htm	2
2	http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	1
2	http://www.swr.de/kaffee-oder-tee/essen/tee/2003/04/16/	1
2	http://www.teebuch.de/teebuch.txt	1
2	http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/	1
2	http://www.teebuch.de/005008.htm	1
2	http://www.pralinen-teeke.de/shop/html/tee.html	1
2	http://www.pandorahs-traum.de/schatztruhe/tee.html	1

B9: Rangfolge der Häufigkeiten von Aufgabe 2 für alle Nutzer

Platz	Rangfolge der relevanten Seiten von Aufgabe 2 für alle Nutzer	Hfgk.
1	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	9
1	http://www.regenwald.org/new/amazonas/highnoon.htm	9
2	http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	5
2	http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	5
2	http://www.der-gruene-faden.de/text/text648.html	5
2	http://www.regenwald-spende.de/ueber_uns.htm	5
3	http://www.econautix.de/site/econautixpage_46.php	4
3	http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms/	4
4	http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	3
4	http://berg.heim.at/tibet/450508/Regen.htm	3
4	http://www.umg.at/112001/abholzung.php	3
4	http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	3
5	http://www.regenwald-institut.de/deutsch/Aktionen/Bolivien.htm	2
5	http://www.abenteuer-regenwald.de/abholz.php	2
5	http://www.umweltlexikon-online.de/fp/archiv/RUBnaturartenschutz/Regenwald.php	2
5	http://www.regenwald.org/new/regenwaldreport/artikel.php?id=147	2
5	http://www.faszination-regenwald.de/	2
5	http://www.cil-frankfurt.de/programme/NEINN/MixMax/Dokumente/CRRegenwald.htm	2
5	http://www.regenwaldschutz.de/	2
6	http://www.regenwald.org/pdf/rdr-report0401.pdf#search='Abholzung%20Regenwald'	1
6	http://www.krref.krefeld.schulen.net/referate/erdkunde/r0235t00.htm	1
6	http://www.hauptschule-gochsheim.de/projekte/reg_auswirk.php	1

B10: Rangfolge der Häufigkeiten von Aufgabe 2 für den Nutzer ROADRUNNERLENNY

Platz	Rangfolge der relevanten Seiten von Aufgabe 2 für ROADRUNNERLENNY	Hfgk.
1	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	3
2	http://www.regenwald.org/new/amazonas/highnoon.htm	2
2	http://berg.heim.at/tibet/450508/Regen.htm	2
3	http://www.abenteuer-regenwald.de/abholz.php	1
3	http://www.regenwald.org/new/regenwaldreport/artikel.php?id=147	1
3	http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	1
3	http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	1
3	http://www.der-gruene-faden.de/text/text648.html	1
3	http://www.faszination-regenwald.de/	1
3	http://www.regenwald-spende.de/ueber_uns.htm	1
3	http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms/	1
3	http://www.regenwald-institut.de/deutsch/Aktionen/Bolivien.htm	1
3	http://www.umg.at/112001/abholzung.php	1
3	http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	1
3	http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	1
3	http://www.regenwaldschutz.de/	1

B11: Rangfolge der Häufigkeiten von Aufgabe 2 für den Nutzer VAH

Platz	Rangfolge der relevanten Seiten von Aufgabe 2 für VAH	Hfgk.
1	http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	2
1	http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	2
1	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	2
1	http://www.regenwald.org/new/amazonas/highnoon.htm	2
1	http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms/	2
1	http://www.regenwald-spende.de/ueber_uns.htm	2
2	http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	1
2	http://www.der-gruene-faden.de/text/text648.html	1
2	http://www.econautix.de/site/econautixpage_46.php	1

B12: Rangfolge der Häufigkeiten von Aufgabe 2 für den Nutzer DASDINGSDA

Platz	Rangfolge der relevanten Seiten von Aufgabe 2 für DASDINGSDA	Hfgk.
1	http://www.regenwald.org/new/amazonas/highnoon.htm	3
1	http://www.der-gruene-faden.de/text/text648.html	3
2	http://www.umweltlexikon-online.de/fp/archiv/RUBnaturartenschutz/Regenwald.php	2
2	http://www.cil-frankfurt.de/programme/NEINN/MixMax/Dokumente/CRRegenwald.htm	2
2	http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	2
3	http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	1
3	http://www.regenwald.org/pdf/rdr-report0401.pdf#search='Abholzung%20Regenwald'	1
3	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	1
3	http://www.regenwaldschutz.de/	1

B13: Rangfolge der Häufigkeiten von Aufgabe 2 für den Nutzer JKESSLER

Platz	Rangfolge der relevanten Seiten von Aufgabe 2 für JKESSLER	Hfgk.
	http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	3
	http://www.econautix.de/site/econautixpage_46.php	2
	http://www.regenwald.org/new/amazonas/highnoon.htm	2
	http://www.regenwald-institut.de/deutsch/Aktionen/Bolivien.htm	1
	http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	1
	http://berg.heim.at/tibet/450508/Regen.htm	1
	http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	1
	http://www.umg.at/112001/abholzung.php	1
	http://www.regenwald-spende.de/ueber_uns.htm	1

B14: Suchergebnisse Google nach Abfrage des Suchbegriffs „Vorzüge von Tee“

------ Ergebnisliste Google für den Suchbegriff "Vorzüge von Tee": -------	Platz b. Testern
1 : http://www.wander.ch/cgi/de/magazine/contact_archive/contact_21/contact1.asp	8
2 : http://www.wander.ch/cgi/de/magazine/contact_current/contact1.asp?print=yes	N/A
3 : http://www.miele.de/D/PR_kva_04.html	N/A
4 : http://www.das-grosse-leben.de/htm/aus%20aller%20Welt.html	N/A
5 : http://teeverband.at/wiss-sort01.htm	1
6 : http://de.chinabroadcast.cn/other/teekultur/23.htm	N/A
7 : http://www.tee-seite.de/aassa.htm	N/A
8 : http://www.milando.de/sq+tee+_167.htm	N/A
9 : https://grubauer.de/product_info.php?products_id=17	N/A
10 : http://www.fm-online.at/jaos/page/main_heute.tmpl?article_id=10010862	N/A
11 : http://www.heuschrecke.com	N/A
12 : http://www.dooyoo.de/wasserfilter/brita-fjord/1021751	N/A
13 : http://www.bauarchiv.de/shopzone24/product.php?asin=B0007WDHZQ&section=4&m	N/A
14 : http://asconet.org:8000/antville/labor	N/A
15 : http://www.preisglocke.de/shop/i_m_naturkosmetik_duschbaeder_+_lotionen_gruener	N/A
16 : http://www.nurnatur.de/gruener_tee.php	N/A
17 : http://www.swr.de/kaffee-oder-tee/essen/tee/2003/04/16	7
18 : http://www.swr.de/kaffee-oder-tee/essen/tee/2003/09/03	N/A
19 : http://www.beategabriele-plus.de/diaetclub/tee/teesorte.html	6
20 : http://www.igv.at/de/id/470/news.details.aspx	N/A
21 : http://www.igv.at/de/id/470/news.details.aspx	N/A
22 : http://www.swr.de/kaffee-oder-tee/essen/tee/2003/09/03	N/A
23 : http://www.teebuch.de/005008.htm	7
24 : http://asconet.org:8000/antville/labor/stories/702	N/A
25 : http://www.freestevia.de/presse/natuerlich_gaertnern.htm	N/A
26 : http://www.shop.energiavital.de/Produkte/Kidneybohnen%20Extrakt/Versandhandel%:	N/A
27 : http://www.shop.energiavital.de/Produkte/Acerola/Kapseln%20Acerola.php	N/A
28 : http://www.shop.energiavital.de/Produkte/Korallenkalzium/Sonderangebote%20Koralle	N/A
29 : http://www.delta-gym.ch/Ernaehrung/nutri/gruener_tee.htm	N/A
30 : http://www.abnehmtreff.de/article133.html	7
31 : http://www.shop.energiavital.de/Produkte/Korallenkalzium/Sonderangebote%20Koralle	N/A
32 : http://www.biocybernetics.de/tonikum.htm	N/A
33 : http://teewalter.de/info35.php	N/A
34 : http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04	2
35 : http://www.swr.de/kaffee-oder-tee/essen/tee/2003/04/16/print.html	N/A
36 : http://www.neuesleben.net/archive/index.shtml	N/A
37 : http://www.faz.net/s/RubD1708818028B4EA7913847E5246256BA/Doc~E327F6949FI	N/A
38 : http://www.wkf.de/newsletter/basis03.html	N/A
39 : http://www.wkf.de/newsletter/basis/basis.pdf	N/A
40 : http://www.bluegreen.net/deutsch/info/texte/stevia.htm	N/A
41 : http://www.golf.at/news/news.asp?id=1145	N/A
42 : http://www.siebenbuerger.de/webshop/Philips-HD-7840/00-Kaffeeautomat-Senseo-Alu	N/A
43 : http://vegetarisch-einkaufen.de/Prod/ProdInf/Milchalt/milchalt.html	N/A
44 : http://www.abnehmtreff.de/article133.html	7
45 : http://lots.bitflux.ch/?start=10	N/A
46 : http://www.factorfake.de/shop-produkt-B0007WDHZQ.html	N/A
47 : http://www.hti.bfh.ch/index.php?id=1891&L=0	N/A
48 : http://www.hti.bfh.ch/index.php?id=1891&L=1	N/A
49 : http://www.hti.bfh.ch/index.php?id=1891&L=2	N/A
50 : http://www.swr.de/imperia/md/content/kaffeeodertee/tee/jasmintee.rtf	N/A

------ Ergebnisliste Google für den Suchbegriff "Vorzüge von Tee": -------	Platz b. Testern
51 : http://de.isodisnatura.ch/nutrition_-_article.htm?ID=14	7
52 : http://www.neuesleben.net/archive/text/gruentee01.shtml	7
53 : http://www.sanoverde.de/index.php/cPath/24_101	N/A
54 : http://www.jede-lebenslage.de/herrenausstatter-13651-olaf_benz_string_instant_jeans	N/A
55 : http://www-x.nzz.ch/folio/archiv/2004/03/articles/turin.html	N/A
56 : http://www.wkf.de/newsletter/0701/wkf0701.doc	N/A
57 : http://fm4.orf.at/janis/196492	N/A
58 : http://www.factorfake.de/shop-produkt-B0007WDHZQ.html	N/A
59 : http://www.palaestinensische-gemeinde.at/gandhi.shtml	N/A
60 : http://www.kapstadt-news.de/news/283.htm	8
61 : http://www.medizin-netz.de/frau/stillenmutter.htm	N/A
62 : http://www.teahouse.de/awards/awards.htm	N/A
63 : http://fm4.orf.at/janis/196492	N/A
64 : http://nopal.aloeshop.de/Produkte/Vitabiosa/Online-Shop%20Vitabiosa.php	N/A
65 : http://www.stachel.de/96.10/10thearix.html	N/A
66 : http://www.wie-gemalt.de/caseking/shop-42510.htm	N/A
67 : http://www.wkf.de/newsletter	N/A
68 : http://www.magic-mount.de/content_magicmount/frame_deutsch.php	N/A
69 : http://www.palaestinensische-gemeinde.at/gandhi.shtml	N/A
70 : http://www.reax.ch/news/news_neuheiten.htm	N/A
71 : http://www.swr.de/kaffee-oder-tee/essen/tee/2005/05/04/print.html	N/A
72 : http://www.medizin-netz.de/frau/stillenmutter.htm	N/A
73 : http://www.teahouse.de/awards/awards.htm	N/A
74 : http://www.parisinfo.de	N/A
75 : http://www.wie-gemalt.de/1apreis/shop-41700.htm	N/A
76 : http://www.kleinanzeigen-landesweit.de/anzeige-51412.html	N/A
77 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/cnid/a953fad62ab33	N/A
78 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/cnid/a953fad62ab33	N/A
79 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/cnid/a953fad62ab33	N/A
80 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/cnid/2f2425eba4b90	N/A
81 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/cnid/2f2425eba4b90	N/A
82 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/cnid/76a4249a9ee8a	N/A
83 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/cnid/19a410d47465d	N/A
84 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/cnid/-/anid/2f2425ea	N/A
85 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/cnid/-/anid/820429b8	N/A
86 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/cnid/-/anid/820429b8	N/A
87 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/cnid/-/anid/6d9426ba	N/A
88 : http://www.jede-lebenslage.de/spezialversandhaus-index-217.html	N/A
89 : http://blog.mellenthin.de/archives/2005/03	N/A
90 : http://www.parisinfo.de	N/A
91 : http://www.logi-methode.de/faq.html	N/A
92 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/fnc/tobasket/cnid/dca	N/A
93 : http://www.callacd.com/oxid.php/sid/x/shp/oxbaseshop/cl/details/fnc/tobasket/cnid/74e	N/A
94 : http://www.sechs-und-sechzig.de/artikel.asp?art=257	N/A
95 : http://www.swr.de/kaffee-oder-tee/vvv/alles-frisch/2003/03/06	N/A
96 : http://shop.meinberlin.de/8381-12S6L1/Mode-Accessoires.html	N/A
97 : http://shopping.lycos.de/search/kaffeevollautomat_espresso.html	N/A
98 : http://www.webvitamine.de/abnehmen/herbal-thermo-stack.html	N/A
99 : http://deutschesfachbuch.de/info/detail.php?isbn=3453141784	7
100 : http://www.wasser.de/aktuell/forum/index.pl?job=thema&tnr=100000000002137&sei	N/A

B15: Suchergebnisse Fooxx nach Abfrage des Suchbegriffs „Vorzüge von Tee“

------ Ergebnisliste Fooxx für den Suchbegriff "Vorzüge von Tee": ------	Platz b. Testern
1 : http://www.tee-special.com/tee.html	1
2 : http://www.pandorahs-traum.de/schatztruhe/tee.html	7
3 : http://www.wendtour.de/ernaehrungtee.htm	8
4 : http://www.herrlein.com/mainpages/teekunderooibuschtee.htm	4
5 : http://www.frauke-spehr.info	6
6 : http://www.steffys.de/htm/gruenertee.htm	8
7 : http://www.teebuch.de/teebuch.txt	5
8 : http://teeverband.at/wiss-sort01.htm	1
9 : http://www.pralinen-teeke.de/shop/html/tee.html	8
10 : http://www.teebuch.de/005008.htm	7
11 : http://www.teeimport.com/teeabc.htm	8
12 : http://www.neuesleben.net/archive/text/gruentee01.shtml	7
13 : http://theology.co.kr/wwwb/CrazyWWWBoard.cgi?db=study&mode=read&n	N/A
14 : http://www.vitalstoff-lexikon.de/index.php?artid=921&mode=showarticle&	3
15 : http://www.geocities.com/Athens/Delphi/3967/tee.html	8
16 : http://www.nurnatur.de/gruener_tee.php	N/A
17 : http://www.henso.com/?idx=157	N/A
18 : http://www.teeblaetter.de/seiten/books.html	N/A
19 : http://www.aalloo.de/essen_trinken_tee_infos.html	N/A
20 : http://www.wkf.de/newsletter	N/A
21 : http://www.atlantis-pharm.com/gtt.htm	N/A
22 : http://www.amazon.de/exec/obidos/ASIN/3517080284	N/A
23 : http://www.nibelungentee.de/tee.pdf	N/A
24 : http://dikigoros.tripod.com/china.htm	N/A
25 : http://asconet.org:8000/antville/labor/stories/702	N/A
26 : http://www.amazon.de/exec/obidos/ASIN/331000497X	N/A
27 : http://www.teeparadies-kailasa.de/Teelander/Tee_in_Europa/tee_in_europa.htm	N/A
28 : http://www.teekoenig.de	N/A
29 : http://www.bezmuri.ch/Trend/pflanzenkunde/Brennesselwirkung.html	N/A
30 : http://home.nikocity.de/jasminfeller/mistel.html	N/A
31 : http://synmed.net/Fachartikel/Mistel.html	N/A
32 : http://www.kolodziej-lieder.de/4639/71602.html	N/A
33 : http://www.tee3000.de/teelink.htm	N/A
34 : http://www.sattwa.de/angebot/tee.html	N/A
35 : http://www.teeparadies-limburg.de/kmgingko.htm	N/A
36 : http://www.teecorner.com/teeinfo.htm	N/A
37 : http://www.demmer.at/teeblatt/ausgabe/fr02/pettigrew.htm	N/A
38 : http://teeladen-nidda.de/01baf193160fe800b/01baf193170f57a0c	N/A
39 : http://www.tag-und-traum.buerger-funk.de/html/teearten.html	N/A
40 : http://www.makrobiotik.com/lexikon.htm	N/A
41 : http://www.tee-kultur.com/pages/tipps.htm	N/A
42 : http://www.geistreich-events.de/blog/creative-environments/328/thinking-carpet-von-v	N/A
43 : http://www.ciao.de/Kombucha__Test_1949420	N/A
44 : http://www.spindel-pro-natur.de/2484.html	N/A
45 : http://www.suedafrikanews.de/news/283.htm	N/A
46 : http://www.ingrids-teeladen.de/shop/produkt180a.htm	N/A
47 : http://igv.at/de/id/470/news.details.aspx	N/A
48 : http://www.fireplanet.de/news/news19.htm	N/A
49 : http://www.dreisesselapotheke.de/htm/ernaehrung/99_10-61_7548_TeeZeit_2_2003.p	7
50 : http://unieasy.ms11.net/shop/k001u003s001.htm	N/A

------ Ergebnisliste Fooxx für den Suchbegriff "Vorzüge von Tee": ------	Platz b. Testern
51 : http://www.miele.de/D/PR_kva_04.html	N/A
52 : http://www.teetrend.de/23001/home.html	N/A
53 : http://www.sabona.tv/Produkte/Schmuck/Power Band/Power Band vergoldet.php	N/A
54 : http://www.sabona.tv/Produkte/Schmuck/Kette/Gold Kette.php	N/A
55 : http://www.wsnetz.de/CHOCODIAET.PDF	N/A
56 : http://www.hexenecke.de/kraeuter.htm	N/A
57 : http://www.andreas-pfab.de/smallt3.htm	N/A
58 : http://www.teeblaetter.de/seiten/books_m.html	N/A
59 : http://energy.pulse.de/life_energy/pflanzenaz2.html	N/A
60 : http://www.zu6.net/lebensmittel/50134894ef0899d04/50134894ef07eac75/50134894ef08	N/A
61 : http://www.shop800.de/familie-kinder/genuss-ernaehrung.kaffee.html	N/A
62 : http://www.de.negoz.com/kollage/resultpage?keyword=tee&stlcmpid=5950	N/A
63 : http://www.stamm-kondor.de/news/zeige.php?was=x5.news	N/A
64 : http://www.tee-shop.net/wellnessprodukte.htm	N/A
65 : http://www.swr.de/kaffee-oder-tee/vvv/alles-frisch/2003/03/06	N/A
66 : http://www.tea-tee.at/teekalender_inhalt.html	N/A
67 : http://www.beepworld.de/members18/hexenkraefte/heilkunde.htm	N/A
68 : http://www.ricola.ch/index.cfm?2AA77C2F2E094E06A546BCC606D3B436	N/A
69 : http://www.magnet-medizin.com/Produkte/Heilen/Verspannungen/Verspannungen Banda	N/A
70 : http://www.magnet-medizin.com/Produkte/Schmuck/Medizinschmuck/Medizinschmuck v	N/A
71 : http://www.reformhaus-kratzert.de/INFO/Ernährung/grünertee.htm	N/A
72 : http://www.teehaus-bierstadt.de/rund_um_die_italienische_pasta_frische_gefuellte_nude	N/A
73 : http://www.globalmedshop.com/i-m-i-m-sensitiv-koerperlotion-gruen...-ml-flasche_det_2(	N/A
74 : http://www.heilig-kreuz.info/html/von_neuenburg_nach_bangladesch5.html	N/A
75 : http://www.swr.de/kaffee-oder-tee/haushaltstipp/2003/03/06	N/A
76 : http://www.yopi.de/Shirts-produktindex__site_1	N/A
77 : http://www.gebueg.de/sengnessel	N/A
78 : http://reformhaus-kurier.de/0411essen_trinken.html	N/A
79 : http://www.teegarten.at/Detail/Teeinfo/Roiboostee.htm	N/A
80 : http://www.juragastroworld.de/cmm/Downloads/files/Download_046.pdf	N/A
81 : http://groups.msn.com/Cafestube/aromatherapie.msnw?pgmarket=de-de	N/A
82 : http://www.megavitalshop.de/shop/getraenke.htm	N/A
83 : http://www.sonias.boutique.ms	N/A
84 : http://groups.msn.com/Cafestube/aromatherapie.msnw	N/A
85 : http://www.perlmeister.com/rundbrief.archiv/20010310	N/A
86 : http://www.welter.de/page-tdt.htm	N/A
87 : http://www.tautropfen.de/scripts/basics/econews/basics.prg?nap=tautropfen	N/A
88 : http://urlaub.azt-ev.de/SriLanka/SriLanka.htm	N/A
89 : http://reformhaus-kurier.de/0108essen_trinken.html	N/A
90 : http://www.vertrieb-und-promotion.com/html/coco_kapseln.html	N/A
91 : http://cydome.com/de/wmiedl/archives/000539.shtml	N/A
92 : http://www.abseits.de/weblog/2003_11_01_archiv.html	N/A
93 : http://asconet.org:8000/antville/labor	N/A
94 : http://www.govindas-naturkost.de/shop1/html/08_ayurveda_tee.htm	N/A
95 : http://webwecker-bielefeld.de/servlet/is/20577	N/A
96 : http://www.hobbythek.de/archiv/294_2	N/A
97 : http://www.hallomahlzeit.de/shop/amorebio/Kaffee__Kakao_und_Tee__...eutel___Ayurv	N/A
98 : http://biorica.info/Deutsch/D-Anti Aging/D-Andere Supplem.htm	N/A
99 : http://www.wellvita-visions.com/daten/04.html	N/A
100 : http://www.traudl-walden.de/Traudi/artikel/publikation_odin.htm	N/A

B16: Suchergebnisse Google nach Abfrage des Suchbegriffs „Abholzung Regenwald“

------ Ergebnisliste Google für den Suchbegriff "Abholzung Regenwald": -------	Platz b. Testern
1 : http://www.uni-koblenz.de/~odsbcg/baeume97/bregenw.htm	1
2 : http://www.regenwald.org/new/amazonas/highnoon.htm	1
3 : http://informationen-hausaufgaben-referat.xn--ad-wia.de/Abholzung-Regenwald-Laecheln-01	N/A
4 : http://informationen-hausaufgaben-referat.xn--ad-wia.de/abholzung-regenwald-weltklima-ihr-(	N/A
5 : http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	4
6 : http://www.omolios.de/omolios68/abholzung_regenwald_gnld.htm	N/A
7 : http://www.energieportal24.de/abholzung%20regenwald/qry_abholzung%20regenwald.htm	N/A
8 : http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	2
9 : http://www.referate-seite.com/referatzimmermann.html	N/A
10 : http://www.regenwald-spende.de/ueber_uns.htm	2
11 : http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	2
12 : http://www.referate-seite.com/referat--weisse-rose.html	N/A
13 : http://www.referate-seite.com/referat-dubai.html	N/A
14 : http://www.referate-seite.com/referate-von-sportstudios.html	N/A
15 : http://www.referate-seite.com/referat-satanismus.html	N/A
16 : http://www.referate-seite.com/referatekleinbildkamera.html	N/A
17 : http://www.der-gruene-faden.de/text/text648.html	2
18 : http://referat-buchvorstellung-hausaufgabe.xn--aa-wia.de/Der-tropische-Regenwald-die-Abh	N/A
19 : http://4zusammenfassung-hausaufgabe-referat.xn--aa-via.de/abholzung-brasilianischer-reg	N/A
20 : http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	4
21 : http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms	3
22 : http://www.referate-seite.com/referat-bundesbeschluss.html	N/A
23 : http://www.referate-seite.com/referate-haus-und-schularbeiten.html	N/A
24 : http://www.referate-seite.com/referat-neutralisation.html	N/A
25 : http://www.referate-seite.com/referat-der-mammut.html	N/A
26 : http://www.referate-seite.com/refferat-ueberleichtatletik.html	N/A
27 : http://www.referate-seite.com/referat-koselleckholmes.html	N/A
28 : http://www.referate-seite.com/refarate.html	N/A
29 : http://www.suchflunder.de/abitur%20erdkunde_60.html	N/A
30 : http://www.4teachers.de/material/2656/Abholzung_im_tropischen_Regenwald.html	N/A
31 : http://www.referate-seite.com/referat-4-klasse.html	N/A
32 : http://www.referate-seite.com/referate-seminar-rechtswissenschaften.html	N/A
33 : http://www.referate-seite.com/referat-ueber-wechselgetriebe.html	N/A
34 : http://www.referate-seite.com/referat-igel.html	N/A
35 : http://www.referate-seite.com/referatirakkrieg.html	N/A
36 : http://www.referate-seite.com/referat-silikonfugen.html	N/A
37 : http://www.referate-seite.com/referate-trigonometri.html	N/A
38 : http://www.referate-seite.com/referat-krankenschwester.html	N/A
39 : http://www.gruene.at/openforum/view.php?site=gruene&bn=gruene_openforum&key=11166	N/A
40 : http://www.umg.at/112001/abholzung.php	4
41 : http://www.referate-seite.com/referat-japan-wirtschaft.html	N/A
42 : http://www.referate-seite.com/referat-bewegungsanalyse-kugelstossen.html	N/A
43 : http://www.referate-seite.com/referat-unternehmenspolitische-entscheidungen.html	N/A
44 : http://www.referate-seite.com/referate-ueber-brasilianische-sportarten.html	N/A
45 : http://www.referate-seite.com/referat-buch-tunnelkids-hausarbeiten.html	N/A
46 : http://www.referate-seite.com/referat-ueber-helmut-newton.html	N/A
47 : http://www.referate-seite.com/referat-1-republik-bis-1934.html	N/A
48 : http://www.referate-seite.com/referat-ueber-rokal-der-steinzeitjaeger.html	N/A
49 : http://www.referate-seite.com/referat-ueber-willhelm-tell.html	N/A
50 : http://blogseek.de/item/20871	N/A

------ Ergebnisliste Google für den Suchbegriff "Abholzung Regenwald": -------	Platz b. Testern
51 : http://www.willemoltmans.nl	N/A
52 : http://www.referate-seite.com/referat-ueber-pferde.html	N/A
53 : http://www.referate-seite.com/referate-freie-wohlfahrtspflege.html	N/A
54 : http://www.referate-seite.com/referat-literaturepoche-realismus.html	N/A
55 : http://www.referate-seite.com/referate-heimat.html	N/A
56 : http://www.referate-seite.com/referat-sylt.html	N/A
57 : http://www.regenwald.org/pdf/rdr-report0401.pdf	N/A
58 : http://www.gruene.at/openforum/view.php?site=gruene&bn=gruene_openforum&key=111(	N/A
59 : http://www.gruene.at/openforum/view.php?site=gruene&bn=gruene_openforum&key=111(	N/A
60 : http://www.gruene.at/openforum/view.php?site=gruene&bn=gruene_openforum&key=111(	N/A
61 : http://www.gruene.at/openforum/view.php?site=gruene&bn=gruene_openforum&key=111(	N/A
62 : http://www.gruene.at/openforum/view.php?site=gruene&bn=gruene_openforum&key=111(	N/A
63 : http://www.referate-seite.com/referate-thueringer-wald.html	N/A
64 : http://www.referate-seite.com/referat-frauenrechte.html	N/A
65 : http://www.referate-seite.com/referat-abendmahl-vinci.html	N/A
66 : http://www.referate-seite.com/referat-ueber-bodypainting.html	N/A
67 : http://www.referate-seite.com/referate-ch.html	N/A
68 : http://www.referate-seite.com/referate-pflege.html	N/A
69 : http://www.referate-seite.com/referat-ueber-weight-watchers.html	N/A
70 : http://www.referate-seite.com/referat-neuengland.html	N/A
71 : http://www.referate-seite.com/referate-ueber-den-2-golfkrieg.html	N/A
72 : http://www.referate-seite.com/referate-bundeswehrreform.html	N/A
73 : http://www.referate-seite.com/referat-logistik.html	N/A
74 : http://www.referate-seite.com/referate-le-corbusier.html	N/A
75 : http://www.referate-seite.com/referat-ueber-tornado.html	N/A
76 : http://www.referate-seite.com/referat-ueber-das-geheimnis-von-bahnsteig-13.html	N/A
77 : http://www.referate-seite.com/referat-stromzaehler.html	N/A
78 : http://www.referate-seite.com/referate-lettland.html	N/A
79 : http://www.regenwald.org/new/regenwaldreport/artikel.php?id=147	5
80 : http://www.awish.net/Europe/germany/photos.html	N/A
81 : http://www.abenteuer-regenwald.de/helfen.php	N/A
82 : http://www.abenteuer-regenwald.de/tiere.php	N/A
83 : http://berg.heim.at/tibet/450508/Regen.htm	4
84 : http://www.autohaus-toeter.de/437271.html	N/A
85 : http://www.autohaus-toeter.de/1152894.html	N/A
86 : http://www.autohaus-toeter.de/2263018.html	N/A
87 : http://www.wuestenfahrer.de/westafrika/westafrika_regenwald.htm	N/A
88 : http://www.parteichef.de/About%20a%20boy	N/A
89 : http://www.fenstervarianten.dasholzalufenster.de	N/A
90 : http://www.econautix.de/site/econautixpage_46.php	3
91 : http://www.autohaus-toeter.de/1447498.html	N/A
92 : http://www.autohaus-toeter.de/2285017.html	N/A
93 : http://www.autohaus-toeter.de/3184595.html	N/A
94 : http://www.autohaus-toeter.de/8310936.html	N/A
95 : http://www.autohaus-toeter.de/4065876.html	N/A
96 : http://www.autohaus-toeter.de/1127583.html	N/A
97 : http://www.autohaus-toeter.de/7154176.html	N/A
98 : http://www.autohaus-toeter.de/7847470.html	N/A
99 : http://www.regenwald.org/new/regenwaldreport/artikel.php?id=119	N/A
100 : http://www.gruene.at/openforum/download_thread.php?site=gruene&bn=gruene_openfor	N/A

B17: Suchergebnisse Fooxx nach Abfrage des Suchbegriffs „Abholzung Regenwald“

------ Ergebnisliste Fooxx für den Suchbegriff "Abholzung Regenwald": -------	Platz b. Testern
1 : http://www.dasgrueneblatt.com/all/regenwald_zerstoerung.php	N/A
2 : http://berlinerforumumweltrecht.de/bin/abholzung_regenwald_69.html	N/A
3 : http://www.wienplan.com/24online/n162/html/162_13_E.html	N/A
4 : http://www.lateinamerika-studien.at/content/lehrgang/lg_mader/lg_mader-212.html	N/A
5 : http://hausarbeit-referat-hausaufgabe.xn--cc-wia.de/Abholzung-im-Regenwald-mich-01-01.l	N/A
6 : http://www.sustentavel.inf.br/arquivos/publica/Oberpfalznetz260804.pdf	N/A
7 : http://www.kbs-koeln.de/webbewerb/hansagym/perukonflikte.html	N/A
8 : http://www.traveldesign.de/tips.html	N/A
9 : http://www.regenwald.org/new/amazonas/highnoon.htm	1
10 : http://www.faszination-regenwald.de/info-center/zerstoerung/abb_28.htm	4
11 : http://www.referate-seite.com/referatzimmermann.html	N/A
12 : http://www.der-gruene-faden.de/text/text648.html	2
13 : http://www.econautix.de/site/econautixpage_46.php	3
14 : http://www.omolios.de/omolios68/abholzung_regenwald_gnld.htm	N/A
15 : http://www.regenwald-spende.de/ueber_uns.htm	2
16 : http://www.abenteuer-regenwald.de/tiere.php	N/A
17 : http://www.ilhacomprida.net/sitio_das_aguas_cantantes.htm	N/A
18 : http://www.regenwaldschutz.de	5
19 : http://www.waldportal.org/books.taiga/print.html	N/A
20 : http://www.umweltfibel.de/lexikon/w/lex_w_wald.htm	4
21 : http://www.brasilien.de/land/florafauna/tropregenwald.asp	N/A
22 : http://www.wuestenfahrer.de/westafrika/westafrika_regenwald.htm	N/A
23 : http://www.umg.at/112001/abholzung.php	4
24 : http://gruppen.greenpeace.de/koblenz/ini2000plus.htm	N/A
25 : http://www.faszination-regenwald.de/info-center/oekosystem/wasserhaushalt.htm	N/A
26 : http://www.regenwald.org/new/regenwaldreport/artikel.php?id=147	5
27 : http://informationen-hausaufgaben-referat.xn--ad-wia.de/Abholzung-Regenwald-Laecheln-	N/A
28 : http://informationen-hausaufgaben-referat.xn--ad-wia.de/abholzung-regenwald-weltklima-i	N/A
29 : http://www.vistaverde.de/news/Natur/0402/12_regenwald.php	2
30 : http://www.energieportal24.de/abholzung regenwald/qry_abholzung regenwald.htm	N/A
31 : http://ccp.ucr.ac.cr/proyecto/poyam/varios/sumariog.htm	N/A
32 : http://www.cil-frankfurt.de/programme/NEINN/MixMax/Dokumente/CRRegenwald.htm	5
33 : http://wwwbestellen.de/index.php/Raubbau	N/A
34 : http://satgeo.zum.de/satgeo/methoden/anwendungen/s506.htm	N/A
35 : http://www.evb.ch/index.cfm?page_id=1391	N/A
36 : http://pharmaka24.de/index.php/ParÃ¡	N/A
37 : http://zahnarzt-krankenversicherung.de/index.php/ParÃ¡	N/A
38 : http://www.kfunigraz.ac.at/communication/news/archiv/2004/helikonien.html	N/A
39 : http://www.spacenight.org/archiv_science_nasa/science_nasa_november2004/16-11-200	N/A
40 : http://referat-buchvorstellung-hausaufgabe.xn--aa-wia.de/Der-tropische-Regenwald-die-Al	N/A
41 : http://4zusammenfassung-hausaufgabe-referat.xn--aa-via.de/abholzung-brasilianischer-re	N/A
42 : http://wcm.krone.at/krone/C00/S15/A7/object_id__30747/hxcms	3
43 : http://www.suchflunder.de/abitur erdkunde_60.html	N/A
44 : http://www.gruene.at/openforum/view.php?site=gruene&bn=gruene_openforum&amp	N/A
45 : http://www.4teachers.de/material/2656/Abholzung_im_tropischen_Regenwald.html	N/A
46 : http://www.vistaverde.de/news/Natur/0201/16_regenwald.htm	N/A
47 : http://www.parlament.ch/afs/data/f/gesch/2001/f_gesch_20013622.htm	N/A
48 : http://prime-forestry.de/en/doc/fscbonn.asp	N/A
49 : http://www.earthlink.de/self.htm	N/A
50 : http://www.greenpeace.de/themen/waelder/nachrichten/artikel/green...rt_mit_amazonas_i	N/A

------ Ergebnisliste Fooxx für den Suchbegriff "Abholzung Regenwald": -------	Platz b. Testern
51 : http://www.greenpeace.de/themen/waelder/nachrichten/artikel/green...rt_mit_amazonas_ir	N/A
52 : http://www.fenstervarianten.dasholzalufenster.de/docs/abholzung_im_regenwald_178.htm	N/A
53 : http://www.gruene.at/openforum/view.php?site=gruene&bn=gruene_openforum&amp	N/A
54 : http://berg.heim.at/tibet/450508/Regen.htm	N/A
55 : http://www.parteichef.de/About a boy	N/A
56 : http://www.regenwald-institut.de/deutsch/Aktionen/Bolivien.htm	N/A
57 : http://www.abenteuer-regenwald.de/helfen.php	N/A
58 : http://berlinerforumumweltrecht.de/bin/regenwald_abholzung_71.html	N/A
59 : http://www.dasholzalufenster.de/docs/abholzung_regenwald_rinder_180.html	N/A
60 : http://www.vistaverde.de/news/Politik/0305/19_regenwald.htm	N/A
61 : http://www.vistaverde.de/news/Politik/0305/19_regenwald.htm	N/A
62 : http://suche-1.de/elektroniki	N/A
63 : http://regenwald-spende.de/ueber_uns.htm	2
64 : http://psychiatrieweb.de/index.php/Ducker	N/A
65 : http://www.fsf.ch/news/Documents/bulletin_105.pdf	N/A
66 : http://www.globales-lernen.de/Schwerpunkte/Regenwald/material	N/A
67 : http://www.regenwald.org/new/newsoftheweek/01-11-26.htm	N/A
68 : http://www.g-o.de/index.php?cmd=focus_detail2&f_id=132&rang=8	N/A
69 : http://www.econautix.de/site/econautixpage_1758.php	N/A
70 : http://www.support-referate.de/geschichte/stalin-lebenslauf-zusammenfassung.htm	N/A
71 : http://www.support-referate.de/geschichte/stalin-lebenslauf-zusammenfassung.htm	N/A
72 : http://www.informatik-im-web.de/referat-gemuese	N/A
73 : http://listi.jpberlin.de/pipermail/attac-konsumnetz/Week-of-Mon-20040913/000524.html	N/A
74 : http://www.sumafox.de/Referat allergie	N/A
75 : http://www.sumafox.de/Referat alzheimer	N/A
76 : http://www.awish.net/Europe/germany/photos.html	N/A
77 : http://www.uvkiel-kai.de/referat-deutsch/Abholzung-Regenwald.html	N/A
78 : http://www.umweltbrief.de/neu/html/archiv/Regenwald.txt	N/A
79 : http://www.globales-lernen.de/Schwerpunkte/Regenwald/material/Samoa.htm	N/A
80 : http://www.wwf.de/young_panda/wissen/Regenwald	N/A
81 : http://www.wwf.de/young_panda/wissen/Regenwald	N/A
82 : http://www.hoteltravel.com/de/malaysia/guides/overview.htm	N/A
83 : http://www.bjrundschau.com/World/2002.43-world-1.htm	N/A
84 : http://www.regenwald.org/new/newsoftheweek/daten2.php?show=84	N/A
85 : http://www.regenzeit.net/de/projekt_regenwald.htm	N/A
86 : http://www.4teachers.de/?action=material&id=2656	N/A
87 : http://www.schnelle-referate.de/beschreibung/interpretation-friedrich-schiller-die.htm	N/A
88 : http://www.schnelle-referate.de/bedeutung/vorteile-nachteile-tintenstrahldrucker.htm	N/A
89 : http://www.informatik-im-web.de/musiksammlungen	N/A
90 : http://www.mirasuch.de/abc+waffen	N/A
91 : http://www.mirasuch.de/abc+waffen	N/A
92 : http://www.referate-im-internet.de/regenwald/sozialversicherungen-bismark.htm	N/A
93 : http://workpage.de/projektw/kape2.php	N/A
94 : http://www.canis.info/oekologie/regenwald_harrypotter.htm	N/A
95 : http://www.ainfos.ca/01/apr/ainfos00239.html	N/A
96 : http://home.eduhi.at/user/loeffler/regenwald.htm	N/A
97 : http://www.regau.gruene.at/ak/regenwald.php	N/A
98 : http://www.regenwald.at/rgs/RWdOE2.html	N/A
99 : http://www.pro-regenwald.org/new_camb.php	N/A
100 : http://www.ebgymhollabrunn.ac.at/projekte/regenwald/konsumverhalten.htm	N/A

B18: Die komplette Personalisierte Ergebnisliste der Nutzer DASDINGSDA und ROADRUNNERLENNY für den Suchbegriff „Vorzüge von Tee“ befinden sich auf der beiliegenden CD unter „Testergebnisse/tee_DD_RR.txt“

B19: Die komplette Personalisierte Ergebnisliste der Nutzer VAH und JKESSLER für den Suchbegriff „Vorzüge von Tee“ befinden sich auf der beiliegenden CD unter „Testergebnisse/tee_VA_JK.txt“

B20: Die komplette Personalisierte Ergebnisliste der Nutzer DASDINGSDA und ROADRUNNERLENNY für den Suchbegriff „Abholzung Regenwald“ befinden sich auf der beiliegenden CD unter „Testergebnisse/regenwald_DD_RR.txt“

B21: Die komplette Personalisierte Ergebnisliste der Nutzer VAH und JKESSLER für den Suchbegriff „Vorzüge von Tee“ befinden sich auf der beiliegenden CD unter „Testergebnisse/regenwald_VA_JK.txt“

B22: Die komplette Personalisierte Ergebnisliste der Nutzer VAH und PKRUG für den Suchbegriff „Apple“ befinden sich auf der beiliegenden CD unter „Testergebnisse/apple_PK_VA.txt“