

Auf der Suche im WWW --- Ein Ansatz zur benutzerorientierten Erstellung von Informationsstrukturen

THOMAS WERNY, HANNOVER
ERNST-WILHELM SCHENK, HANNOVER

Abstract

This paper presents a prototype of an information system, which will search in a well defined domain for user relevant informations using user profiles. The returned informations can then be rated by the users to update or improve their profiles. It is also possible to group a couple of user ratings to form a group profile

1 Einführung

Seit der Einführung des ersten Web-Browsers im Jahre 1993 hat das World Wide Web eine rasante Entwicklung erfahren. Während die Anzahl der WWW-Server im Dezember 1994 noch ca. 10.000 betrug wird die aktuelle Zahl von ca. 650.000 auf bis zu 2 Millionen WWW-Server geschätzt (siehe hierzu [1] und [2]). Durch den großen Umfang wird es heutzutage aber immer schwieriger die 'richtigen' Informationen, d.h. relevante und aktuelle Informationen im WWW zu finden. In immer kürzeren Zeitabständen kommen neue Informationen hinzu, so daß die Menge der vorhandenen Informationen für den Informationssuchenden oft nicht mehr überschaut werden kann (siehe [RRZN98]). Auch die zur Zeit für die Suche im WWW vorhandenen Suchhilfen reichen nicht mehr aus, da sie teilweise veraltete bzw. nicht mehr gültige sowie irrelevante aber auch zugleich unvollständige Informationen liefern. Das bedeutet, daß man eine große Menge von Verweisen bekommt, aus denen der Benutzer anschließend selbst die wenigen relevanten Informationen extrahieren muß. Die verschiedenen Suchhilfen dienen dabei nur als eine Art Vorfilter, um die 'grobe' Sucharbeit zu erledigen. Die 'feine' Sucharbeit muß der Benutzer selbst durchführen. Oft kommt hier noch hinzu, daß die vorhandenen Informationen unzureichend bis überhaupt nicht strukturiert sind. Der einzelne Endbenutzer, der nur an den reinen Informationen interessiert ist und nicht an deren Beschaffung, hat nicht die Zeit und teilweise auch nicht das notwendige Wissen, um die für ihn relevanten Informationen zu finden.

Bestehende Systeme behandeln stets nur Teile dieser Problematik. So unterstützen z. B. Suchmaschinen (siehe [SRR97]), die über umfangreiche Indizes verfügen, immer komplexere Anfragen, die aber zumeist nur auf das Vorhandensein/Nichtvorhandensein einer Menge von Begriffen prüfen. Systeme, die dagegen einen inhaltlichen Bezug verwenden, betrachten oft nur eine sehr eingeschränkte Menge von WWW-Seiten oder reduzieren die Auswahl auf Seiten mit einer bekannten Struktur, z. B. Zeitungen (siehe [BILC97] und [KBA95]).

2 Zielsetzung

Es wird hier ein Informationssystem vorgestellt, das den Endbenutzer bei der Suche nach für ihn relevanten Informationen unterstützt bzw. ihm diese teilweise oder ganz abnimmt (in diesem Fall: die Suche nach HTML-Seiten). Der Endbenutzer fragt nur noch die zu einem für ihn erstellten Benutzerprofil gesammelten Informationen ab und bekommt sofort die im Informationssystem gespeicherten aktuell vorhandenen Informationen angezeigt. Darüberhinaus kann

der Benutzer eine Bewertung des zurückgelieferten Ergebnisses vornehmen und es existiert eine Gruppensicht. Das hier vorgestellte Informationssystem wurde für das Fachgebiet Gartenbau realisiert.

3 Realisierung

Eine Skizze des realisierten Informationssystems, welche die verschiedenen Abläufe und Zusammenhänge visualisiert, ist in Abbildung 1 dargestellt. Die wichtigsten Komponenten sind zum einen der Experte, die verschiedenen Benutzer zusammen mit ihren Benutzerprofilen und auf der anderen Seite die Gatherer- und die Indexkomponente. Nachfolgend werden die einzelnen Komponenten genauer beschrieben. Die durchgezogenen Pfeile stellen hierbei Aktionen dar, die ausgeführt werden, die unterbrochenen Pfeile stellen die Beeinflussung von verschiedenen Aktionen durch das Benutzerprofil dar.

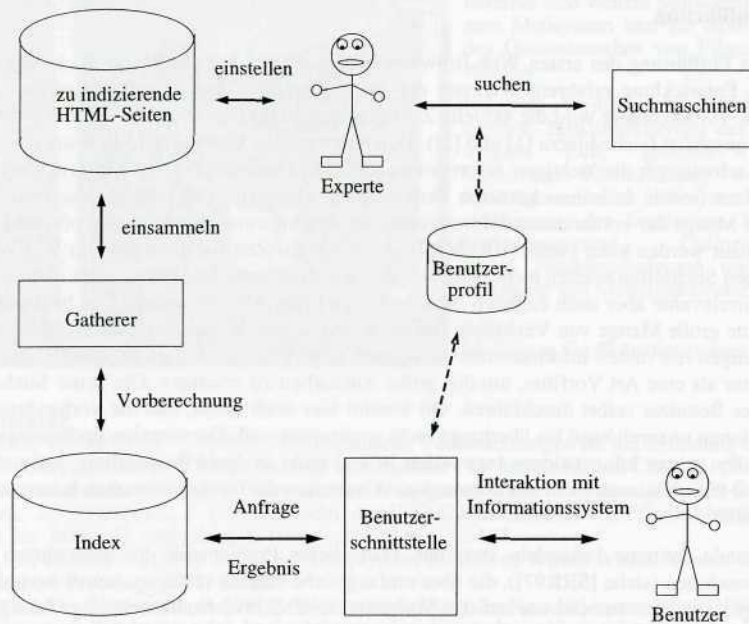


Abbildung 1: Aufbau des Informationssystems

3.1 Der Experte

Der Experte legt eine Sammlung von Verweisen ab. Diese Verweise entsprechen den WWW-Servern bzw. Teilen der WWW-Server, die für den Gartenbau wichtige Informationen enthalten. Zur Ermittlung dieser Verweise kann er selbst eine Suche in verschiedenen Such- und Metasuchmaschinen durchführen oder bekommt interessante Verweise von Anwendern übermittelt. Diese neuen Verweise werden dann von ihm auf ihre Relevanz überprüft und gegebenenfalls in die Verweissammlung eingestellt. Auf diese Weise ist gewährleistet, daß keine thematisch irrelevanten HTML-Seiten behandelt werden.

3.2 Gatherer und Index

Die vom Experten eingestellten Verweise werden von dem Gatherer bis zu einer vorgegebenen Tiefe eingesammelt. Anschließend wird mit der Indexkomponente ein Index erstellt. In diesen Index geht die Struktur der HTML-Seite mit ein, d.h. es wird z.B. festgehalten, ob ein Begriff im Titel oder in einer Aufzählung auftritt. Dadurch ist es möglich Anfragen, die sich auf die Struktur einer HTML-Seite beziehen, zu realisieren.

3.3 Der Benutzer, Benutzerprofile und Bewertung

Benutzerprofile beschreiben die Interessen des einzelnen Benutzers. Entsprechend dieser Benutzerprofile werden alle Informationen aufbereitet und den Benutzern zur Verfügung gestellt.

Der Benutzer kann eine Bewertung der als Ergebnis zurückgelieferten HTML-Seiten vornehmen („gut“ oder „schlecht“). Eine Bewertung hat zur Folge, daß das Informationssystem anhand der bewerteten HTML-Seiten überprüft, ob dem Benutzer ein neues Benutzerprofil, das besser zu den bewerteten HTML-Seiten paßt als das vorherige, vorgeschlagen werden kann. Der Benutzer kann diesen Vorschlag akzeptieren oder ihn verwerfen. Der Vorschlag wird durch die Überprüfung der Struktur der bewerteten HTML-Seiten generiert.

In Abbildung 2 ist das Zusammenspiel von Benutzer und Benutzerprofil dargestellt. Benutzer können im einfachsten Fall nur eine Anfrage stellen und sich das zurückgelieferte Ergebnis ansehen. Darüberhinaus kann ein Benutzer eine Anfrage als Profil abspeichern, sein Profil ändern, sein Profil abfragen, d.h. alle Informationen anzeigen lassen, die zu diesem Profil passen oder ein zurückgeliefertes Ergebnis bewerten.

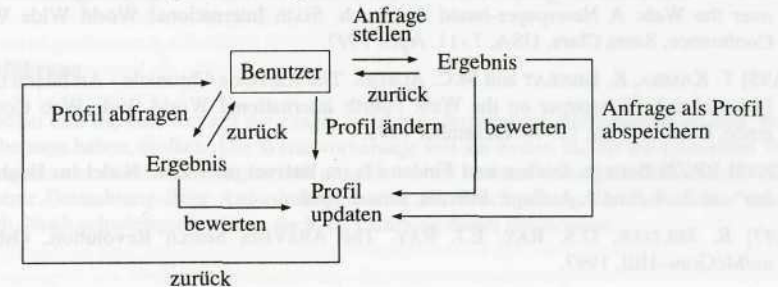


Abbildung 2: Zusammenspiel Benutzer und Profil

3.4 Gruppenprofil

Sind die verschiedenen Benutzer einzelnen Gruppen zugeordnet, dann kann man die einzelnen Benutzerbewertungen auch für eine Gruppenbewertung nutzen. Das Gruppenprofil bietet die Möglichkeit die bewerteten HTML-Seiten von Benutzern, die der gleichen Gruppe zugeordnet sind, zusammenzufassen und sowohl den dem Informationssystem bekannten Benutzern als auch fremden Benutzern als Orientierungshilfe zur Verfügung zu stellen. Dem Informationssystem bekannte Benutzer können dann diese Gruppeninformationen ebenfalls wieder bewerten.

4 Zusammenfassung

Das vorgestellte Informationssystem kann für einen gesamten Bereich, z.B. den Gartenbau oder auch wieder für ein Teilgebiet eines Bereiches eingesetzt werden, z.B. das Teilgebiet Pflanzenkrankheiten. Dem Experten kommt hierbei eine wichtige Funktion zu, da er die WWW-Server oder Ausschnitte (Verzeichnisse oder Dateien) davon zusammenstellt, die für ein Teilgebiet relevant sind. Ausgehend von dieser Zusammenstellung werden von den verschiedenen WWW-Server die relevanten Informationen eingesammelt. Aus den eingesammelten Informationen werden dann die für die einzelnen Benutzer relevanten Informationen, entsprechend dem Benutzerprofil, extrahiert, so daß diese nur den Teil der Informationen aus dem WWW bekommen, der sie interessiert. Mit den hierbei gefundenen Informationen werden anschließend verschiedene 'Weiterverarbeitungsschritte' durchgeführt bevor sie dem Benutzer präsentiert werden. Diese 'Weiterverarbeitungsschritte' umfassen z.B. eine Analyse der HTML-Struktur und eine Bewertung der Informationen, mit deren Hilfe schließlich eine benutzergerechte Aufarbeitung und Präsentation der Suchergebnisse erzielt wird. Durch die Möglichkeit, zurückgelieferte HTML-Seiten zu bewerten, kann der Benutzer gezielt auf die folgenden Suchen Einfluß nehmen bzw. es können gezielt Gruppeninformationen, d.h. bewertete Informationen von Benutzern, die in der gleichen Gruppe eingetragen sind, erstellt werden.

5 Literatur

[BILC97] M. BANÂTRE, V. ISSARNY, F. LELEU, B. CHARPIOT. Providing Quality of Service over the Web: A Newspaper-based Approach. Sixth International World Wide Web Conference, Santa Clara, USA, 7.-11. April 1997.

[KBA95] T. KAMBA, K. BHARAT and M.C. ALBERS. The Karkatoa Chronicle - An Interactive, Personalized, Newspaper on the Web. Fourth International World Wide Web Conference, Boston, USA, 11.-14. Dezember 1995.

[RRZN98] RRZN-Beilage. Suchen und Finden (?) im Internet oder „Die Nadel im Heuhaufen“ zur 3., 4., und 5. Auflage: Internet, Januar 1998.

[SRR97] R. SELTZER, D.S. RAY, E.J. RAY. The AltaVista Search Revolution. Osborne/McGraw-Hill, 1997.

[1] <http://www.nw.com/zone/WWW/firstnames.html>

[2] <http://www.netcraft.com/Survey/>