

Effiziente und proaktive Suche

in Helpdesk Anfragen mit CompleteSearch

Holger Bast, Ingmar Weber

CompleteSearch ist eine interaktive Suchmaschine, die den Nutzer, während er eine Suchanfrage tippt, nach jedem Tastendruck mit Feedback unterstützt, wie z.B. Vorschlägen für Vervollständigung der Suchanfrage oder eine Kategorisierung der bisherigen Treffer. Durch diese proaktive und interaktive Natur des Systems wird es für den Nutzer leichter, relevante Dokumente in einer Sammlung von Erfahrungswissen zu finden. Wir haben den Datensatz des Helpdesks unseres Institutes (eine Sammlung von rund 8,000 Hilfsanfragen mit Lösungsverlauf) mit CompleteSearch indiziert und für die Helpdeskmitarbeiter durchsuchbar gemacht. Im Rahmen einer kleinen im Oktober 2006 durchgeführten Nutzerstudie zeigen wir, dass CompleteSearch im Vergleich zu einer herkömmlichen Volltextsuchmaschine (Google Desktop Search) klare Vorteile bietet: Es werden mehr relevante Dokumente gefunden und alle Testnutzer bevorzugten CompleteSearch wegen seiner Geschwindigkeit und des Gefühls durch das System "bemächtigt" zu werden. Wir berichten auch über einige Erkenntnisse, die wir im Zusammenhang mit der Studie gewonnen haben und die für die Planung ähnlicher Systeme interessant sein sollten.

1 Introduction

Kurze Antwortzeiten sind essentiell wichtig, wenn es darum geht, Nutzer in ihrer Suche *interaktiv* zu unterstützen. Das CompleteSearch System [1] wurde genau mit der Motivation entwickelt, eine Reihe von "intelligenten" Suchfähigkeiten auch auf großen Textsammlungen interaktiv (nach jedem Tastendruck) bieten zu können.

Um den praktischen Nutzen unseres Systems zu demonstrieren, haben wir den Datensatz des Helpdesks unseres Institutes mit CompleteSearch indiziert und eine kleine Nutzerstudie mit fünf Helpdeskmitarbeitern durchgeführt. Die Daten des Helpdesks enthalten rund 8,000 "Tasks", d.h. Dokumente mit Hilfsanfragen von einigen hundert Nutzern sowie das anschließende Bearbeitungsprotokoll mit Emailkorrespondenz zwischen Nutzern und Helpdesk.

Für unsere Studie baten wir die fünf Teilnehmer, relevante Dokumente für insgesamt 10 fiktive Probleme zu finden. Dafür nutzten sie teils unser CompleteSearch System und teils Google Desktop Search (desktop.google.com). Alle fünf Teilnehmer bevorzugten CompleteSearch gegenüber Google Desktop, hauptsächlich wegen seiner Geschwindigkeit, des Gefühls "die Kontrolle zu haben" und der erweiterten Suchmöglichkeiten.

In Kapitel 2 beschreiben wir das CompleteSearch System im Detail. In Kapitel 3 betrachten wir relevante andere Arbeiten. In Kapitel 4 beschreiben wir die Durchführung der Nutzerstudie und ihre Ergebnisse. Die vorliegende Arbeit ist in weiten Teilen eine deutschsprachige Überarbeitung der Ergebnisse aus [2].

Aus Datenschutzgründen können wir leider keinen Demonstrator des CompleteSearch Systems für den Helpdeskdatensatz anbieten. Stattdessen bieten wir unter <http://search.mpi-inf.mpg.de/library> einen ähnlichen Demonstrator zum Durchsuchen einer großen Informatikbibliothek an.

2 Das CompleteSearch System

Der Kern des CompleteSearch Systems besteht aus einem Mechanismus, der Instanzen des folgenden *Präfixsuch- und Vervollständigungsproblems* effizient löst:

Für eine unvollständig getippte Suchanfrage berechne (i) die Vervollständigungen des letzten (nur teilweise getippten) Suchwortes, die zu den besten Treffern führen, so wie (ii) eben diese besten Trefferdokumente. Der Screenshot in Abbildung 1 zeigt ein Beispiel. Eine formale Definition des algorithmischen Problems, eine Variante eines mehrdimensionalen Bereichssuchproblems, findet sich in [1].

Auf den ersten Blick erscheinen die Möglichkeiten solch eines Mechanismus als rein syntaktisch: es geht um Worte, die mit einer bestimmten Buchstabenfolge beginnen. Im Folgenden zeigen wir, anhand einiger Beispiele, wie sich durch die Ergänzung geeigneter Kunstworte zum Index eine Vielzahl von mehr und mehr semantischen Fähigkeiten realisieren lässt. Dieser Ansatz entspricht im Wesentlichen der im Case-Based Reasoning üblichen Idee, zusätzliches Wissen durch ein kontrolliertes Vokabular zu repräsentieren.

Semantische Annotationen. Sei das Wort `kmhp81001` durch beliebige Annotationshilfsmittel als Druckernamen (`printer`) identifiziert. Wir fügen dann die Worte `printer:kmhp81001` und `kmhp81001:printer` den entsprechenden Dokumenten und damit dem Index hinzu. Falls ein Nutzer dann `problem pri` tippt, so könnte eine mögliche Vervollständigung `printer:kmhp81001` sein (was von unserem Userinterface als `kmhp81001`, `the PRINTER` dargestellt wird), vorausgesetzt es gibt mindestens ein Dokument das sowohl `problem` als auch `kmhp81001` enthält (bzw. Worte, die mit diesen Präfixen beginnen). Ähnlich geht man für die Anfrage `problem kmhp` vor.

Facettensuche. Nehmen wir an, es ist bekannt, dass ein Dokument zu bestimmten Kategorien gehört, z.B. `Linux` und `Mail`. Dann fügen wir die Worte `cat:linux` und `cat:mail` unserem System hinzu, d.h., sie werden in den entsprechenden Dokumenten künstlich ergänzt. Für die Suchanfrage `network`