



GETESS - Intelligent Suchen in Datennetzen

Keywords: NLP, IR, Dialogsystem, WWW, Data Mining, Datenbanken, XML, HTML, natürliche Sprache, Ontologie

Informationsrecherche im Internet beschränkt sich heute weitestgehend auf IR-Suchtechniken, meist jedoch auf Stichwortsuche unter Nutzung von booleschen Operatoren. Zusätzlich werden hierarchisch strukturierte Themengebiete angeboten. Auch wenn Suchanfragen an google (www.google.com) erstaunlich oft gute Resultate liefern, für die meisten Benutzer ist die Formulierung einer exakten Suchanfrage zu ihrem Problem viel zu kompliziert und bei unscharfen Formulierungen versagen die Suchdienste kläglich.

Dabei verfügt der Mensch über ein starkes Ausdrucksmittel - **die natürliche Sprache**.

Vor diesem Hintergrund haben sich 1998 vier Partner aus Forschung und Industrie zusammengefunden und das Projekt **GETESS** gestartet, das vom BMBF gefördert wurde. Das Ziel des Projektes war es, ein intelligentes Werkzeug zur Informationsrecherche im Internet zu entwickeln. Dabei kommen Techniken und Verfahren aus unterschiedlichen Gebieten der Informatik zum Einsatz. Die folgenden Anforderungen wurden formuliert:

- * das System soll einfach zu bedienen sein.
- * zentrale Wissensbasis ist eine Ontologie.
- * automatisches Einsammeln der Informationen aus Text- und HTML-Dateien (deutsch/englisch) und Abspeichern in einer Datenbank.
- * Anfragen an das System können in natürlicher Sprache (deutsch/englisch) gestellt werden. GETESS behandelt Probleme bei der Anfrage im Dialog mit dem Benutzer
- * moderne Multimediatechniken kommen zum Einsatz.
- * Einschränkung auf spezielle Themengebiete (Domains).

Die jeweiligen Projektpartner bringen umfangreiches Wissen und Erfahrungen auf ihrem Spezialgebiet in das Projekt mit ein:

- * DFKI Saarbrücken - Analyse von natürlichsprachlichem Text (parsen, generieren)
- * AIFB Uni Karlsruhe - Ontologien als zentrale Wissensbasen
- * DBIS Uni Rostock - Datenbanktechnologien, IR-Systeme
- * GECKO mbH Rostock - Projektmanagement, Dialogsystem

Die Idee von GETESS besteht nun darin, dass bei der syntaktischen Analyse eines Dokumentes, erkannte Begriffe und Beziehungen semantisch mit Informationen aus der Ontologie angereichert werden. Kann zwischen zwei erkannten Begriffen eine Beziehung hergestellt werden, wird ein Abstrakt generiert und abgespeichert. Eine Menge von Abstrakts beschreibt dann ein Dokument.

Werden die selben Techniken der linguistischen und semantischen Analyse auch auf der Benutzerseite verwendet, können die Anfragen mit den gespeicherten Abstrakts verglichen werden. Als Ergebnis wird dann eine (geordnete) Menge von gefundenen Dokumenten präsentiert.

Treten bei den Bearbeitung der Anfrage Probleme auf (etwa es werden eingegebene Begriffe nicht erkannt oder keine bzw. zu viele Ergebnisse gefunden) initiiert das System einen Klärungsdialog.

Nach dreijähriger intensiver Arbeit kann ein Prototyp präsentiert werden, der alle gestellten Anforderungen erfüllt und damit die Machbarkeit der im Antrag formulierten Ziele zeigt.