

Pressemitteilung

Universität Osnabrück

Marietta Fuhrmann-Koch

22.07.1998

<http://idw-online.de/de/news5833>

Forschungsprojekte
Informationstechnik, Medien- und Kommunikationswissenschaften, Sprache / Literatur
überregional

Recherche nach Literatur: OSIRIS sucht mit "Verstand"

Die Universitätsbibliothek Osnabrück hat zusammen mit dem Institut für Semantische Informationsverarbeitung der Universität Osnabrück in zweijähriger Arbeit eine computerbasierte intelligente Recherchemöglichkeit für Literatur entwickelt: Das System OSIRIS, das bereits in der Praxis eingesetzt wird, kann die Sucheingabe eines Benutzers "verstehen". Das Projekt wird von der Deutschen Forschungsgemeinschaft (DFG) gefördert.

Pressemitteilung
Osnabrück, 22. Juli 1998 / Nr. 114/98

Recherche nach Literatur: OSIRIS sucht mit "Verstand"
Universitätsbibliothek und Forschungsinstitut entwickelten Intelligentes Suchsystem

Wer heute Literatur für Forschung, Lehre oder Studium sucht, muß längst nicht mehr Karteikarte für Karteikarte im Zettelkasten durchforsten, sondern kann sich auf Computerhilfe stützen. In der Universitätsbibliothek Osnabrück ist seit 1993 ein Online Public Access Catalog (OPAC) im Einsatz, der jedoch "wie alle konventionellen Online-Benutzerkataloge deutliche Schwächen bei der Büchersuche zeigt", so der stellvertretende Bibliotheksdirektor und EDV-Experte Dr. Hartmut Zillmann. In zweijähriger Arbeit hat die Universitätsbibliothek nun zusammen mit dem Institut für Semantische Informationsverarbeitung der Universität Osnabrück eine intelligente Recherchemöglichkeit entwickelt: Das System OSIRIS, das bereits in der Praxis eingesetzt wird, kann die Sucheingabe eines Benutzers im wahrsten Sinne des Wortes "verstehen". Das Projekt wird von der Deutschen Forschungsgemeinschaft (DFG) gefördert.

Jeder Bibliotheksbenutzer ist vermutlich schon einmal in diese Situation geraten: Der Computer soll Publikationen zu einem Thema heraussuchen, zum Beispiel zum Sport mit älteren Menschen. Bei der Stichworteingabe Sport stößt das System an seine Grenzen. Die Trefferquote liegt bei über 1.000 Titeln. Der OPAC zeigt an, daß die einzelnen Werke nicht mehr ausgewiesen werden können. Die Kombination Alterssport oder Seniorensport fördert zwar einige Bücher zutage, aber aus dem Ergebnis ist nur schwer ersichtlich, ob dies wirklich alle Werke zum Thema sind. Der dritte Versuch - Sport mit älteren Menschen - scheitert ganz. "Das Problem liegt in der Art, wie im OPAC nach der Literatur gefahndet wird", erläutert Marc Ronthaler vom Institut für Semantische Informationsverarbeitung.

Grundsätzlich unterscheidet sich das System dabei kaum vom Zettelkasten: Beiden liegt, so der Computerlinguist Ronthaler, eine Sammlung von Datensätzen zugrunde, die ausgewertet werden müssen. Hat der Benutzer der Karteikarten noch jede einzelne Karte für sich bewertet, so erfolgt die Suche im OPAC rein mechanisch: Das eingegebene Titelstichwort - zusammen mit der Option Personennamen die am häufigsten angewandte Suchmethode - wird mit den vorhandenen Daten abgeglichen. Das System listet alle Werke auf, die dieses Stichwort im Titel enthalten. Marc Ronthaler: "Bei zu allgemein gefaßten Stichwörtern kommt es so zu riesigen Treffermengen."

Der umgekehrte Fall hat ähnlich ärgerliche Konsequenzen: Nur wer mit dem Stichwort genau die Titelformulierung trifft, hat bei der Suche Erfolg. Die Suchanfrage italienischer Film findet weder das Lexikon des italienischen Films noch die

Geschichte italienischer Filme. "Darüber hinaus sind die Datensätze im OPAC inhaltlich äußerst dürftig deskribiert", bemängelt Dr. Zillmann. Ergänzende Angaben, angefügt durch den Bibliothekar, bilden "die Ausnahme". So sei es schon Zufall, wenn ein Suchbegriff mit der Erfassung im OPAC übereinstimme. Und schließlich kritisiert Dr. Zillmann "die mangelnde Flexibilität des Systems gegenüber fehlerhaften Benutzereingaben", die ebenfalls zu Null-Treffer-Ergebnissen führe.

Auch Lösungsstrategien wie Schlagwörter und Trunkierung stoßen nach den Worten der beiden Osnabrücker Experten "schnell an ihre Grenzen". Die personalintensive Erschließung durch Schlagwörter verursacht hohe Kosten. Daher ist nur etwa die Hälfte der rund eine Million Bücher in der Universitätsbibliothek verschlagwortet, so Dr. Zillmann. Die Trunkierung - das Abschneiden eines Wortes und die Verwendung eines Platzhalters - dagegen führt vielfach zu unpräzisen Ergebnissen. Die Suche nach Auto* zeigt nicht nur Autodach oder Autotür, sondern ebenso Automatisierung. Und auch für semantische Phänomene stellt sie keine Lösung dar: Der Begriff Morphologie wird mit oder ohne Platzhalter quer durch die Disziplinen Biologie, Geologie, Linguistik und Medizin erfaßt.

Auf diese Probleme reagiert das Osnabrück Intelligent Research Information System, kurz OSIRIS. Mit Hilfe computerlinguistischer Komponenten analysiert OSIRIS die Suchanfragen sowohl syntaktisch als auch semantisch. Dr. Zillmann: "Das System versteht - mit Einschränkungen - die deutsche Schriftsprache." Der Benutzer kann daher Anfragen weitgehend so eingeben, wie er sie einem Menschen gegenüber ausdrücken würde. OSIRIS unterscheidet, ob nach Wald im Unterricht oder Unterricht im Wald gefragt wird. Außerdem hilft das System bei der notorisch schwierigen Schreibung von Namen. Im Gegensatz zum OPAC zeigt OSIRIS auch Suchergebnisse bei Isabelle Ebers an, obwohl der Name korrekt Isabella Schneider-Eberz lautet. Möglich macht dies der Einsatz spezieller phonetischer Algorithmen.

Auch durch einen zweiten Aspekt trägt OSIRIS zu einer qualitativen Verbesserung der Recherche bei: Das Research System ist in der Lage, den verfügbaren Datenbestand "intelligent aufzubereiten", betonen die beiden OSIRIS-Väter Zillmann und Ronthaler. Das bedeutet: Aus dem OPAC sammelt und bewertet OSIRIS all die Informationen, die der Benutzer normalerweise nicht sieht. Dieser automatisch generierten Wissensbasis werden später regelmäßig weitere Daten über neuerworbene Werke hinzugefügt.

Seit Herbst 1997 steht OSIRIS - inzwischen mit der Version 2.1 - den Mitgliedern der Universität zur Verfügung. Seither wird das System am Fachbereich Psychologie unter der Leitung von Dr. Kai-Christoph Hamborg evaluiert. Ziel ist es, Benutzeroberfläche und Funktionalität weiter zu verbessern. Die DFG hat die Förderung des Projektes OSIRIS inzwischen bis 1999 verlängert. In Übereinstimmung mit den Statuten der Deutschen Forschungsgemeinschaft ist für die bei OSIRIS verwendete Technologie nun auch eine wirtschaftliche Nutzung vorgesehen.

Kontaktadresse:
Dr. Hartmut Zillmann
Universitätsbibliothek Osnabrück
Alte Münze 16, 49034 Osnabrück
Tel. (0541) 969-4359, Fax (0541) 969-4482
E-Mail: hartmut.zillmann@ub.uni-osnabrueck.de

URL zur Pressemitteilung: <http://www.ub.uni-osnabrueck.de/>