

Cross-Language Information Retrieval und automatische Sacherschließung in Suchmaschinen am Beispiel der „Bielefeld Academic Search Engine“ (BASE)

Referent: Dirk Pieper, Bielefeld

Abstract:

Der Begriff World-Wide-Web impliziert einen globalen Zugang zu Informations-Ressourcen, faktisch existieren für viele Nutzer je nach Sprachhintergrund verschiedene Teilausschnitte: ein English-Wide-Web, ein Deutsch-Wide-Web usw. Cross-Language Information Retrieval (CLIR) versucht, das Auffinden von anderssprachigen Dokumenten zu erleichtern, in dem in einer bestimmten Sprache formulierte Anfragen automatisch entsprechende anderssprachige Dokumente finden.

Eine Möglichkeit der Realisierung von CLIR ist der Einsatz von multilingualen Thesauri. Im Rahmen der Weiterentwicklung der „Bielefeld Academic Search Engine“ (BASE) wurde mit EUROVOC ein solcher Thesaurus implementiert. Der vom „Amt für amtliche Veröffentlichungen der Europäischen Gemeinschaften“ in Luxemburg gepflegte Thesaurus umfasst 21 Sprachen, darüber hinaus gibt es die Option, zusätzlich – sofern im Thesaurus vorhanden – synonyme Begriffe zu finden.

Der Vortrag beschreibt, wie OAI-Metadaten mit angereicherten Thesaurus-Daten von Suchmaschinen indexiert werden können. Mit einer ähnlichen Technik ist es dann auch möglich, weitere Wörterbücher oder Thesauri in den Indexierungsprozess einzubinden, was am Beispiel der Schlagwort-Normdatei (SWD) demonstriert wird.