

2.5 Konzeptionelle Anfrage-Verfeinerung

Das Hauptproblem klassischer Methoden für die Anfrage-Verfeinerung besteht darin, dass Anfragen als „bag of words“ betrachtet werden, so dass die erzeugten Verfeinerungen keine semantische Verbindung mit der Anfrage haben (d.h. der Verfeinerungs-Prozess wird auf syntaktischer Ebene durchgeführt).

Im Rahmen des Projekts wurde eine neue Methode, die sogenannte Konzeptionelle Anfrage-Verfeinerung, entwickelt, die mit Hilfe von Ontologien versucht, die semantische Bedeutung einer Anfrage nachzuvollziehen. Die Methode basiert auf dem Librarian Agent Query Refinement Process. Das ist ein drei-stufiger Prozess bestehend aus: 1. der Messung der Mehrdeutigkeiten einer Anfrage, 2. der Empfehlung der Verfeinerungen der Anfrage und 3. dem Ranking von Verfeinerungen.

Die Aufgabe des Verfeinerungs-Prozesses ist es, die Verfeinerungen, die zu der semantischen Bedeutung der Anfrage passen, zu generieren. D.h. eine Verfeinerung sollte eine konkrete semantische Rolle bezüglich der Anfrage spielen. Nach der Informationstheorie gibt es zwei generelle Relationen, welche die Bedeutung eines Begriffes erläutern können: Specialisation und Modification. Diese semantischen Relationen zwischen den Begriffen der Anfrage und den Begriffen des Textes zu produzieren, ist das Ziel der entwickelten Methode. Auf diese Weise sind die Verfeinerungen, die diese Methode generiert, nicht nur zusätzliche Begriffe in einer Anfrage, sondern semantische Erweiterungen der Anfrage. So ist z.B. für die Anfrage „code“ der Spezialisierungs-Begriff „binary“ (siehe oben) eine semantische Erweiterung: „code + binary“ bedeutet „binary code“ (binärer Code) und nicht eine beliebige Relation zwischen „code“ and „binary“. Infolgedessen produziert unsere Methode Verfeinerungen, die relevant für die Benutzer-Präferenzen sind.

Mit Hilfe von Natural Language Processing (NLP) kann man die Relationen Specialisation und Modification in einem Text entdecken. Weil NLP-Methoden oft sehr zeitintensiv sind, wurde eine neue Methode entwickelt, die flaches Tagging mit logischem Inferenzing kombiniert. Auf diese Weise wird von der linguistischen Verarbeitung (d.h. wir benutzen linguistische Strukturen, wie z.B. noun phrases) und vom logischen Inferenzing (d.h. mit Hilfe von Regeln produzieren wir neue Informationen, die für die Verfeinerungen wichtig sind) profitiert.

Die Hauptvorteile für einen Benutzer sind:

- a) bessere Qualität des Suchprozesses: kürzere Suchzeit, bessere Exploration des Suchraums und bessere Berücksichtigung der Präferenzen des Benutzers sowie
- b) bessere Qualität der Ergebnisse: die Ergebnisse sind geclustert und relevanter.