
Abstract

Die vorliegende Diplomarbeit „Entwicklung einer Suchmaschine für wissenschaftliche Dokumente“ beschreibt den Entwurf und die Entwicklung einer Volltextsuchmaschine, die im Rahmen eines Projektes beim Deutschen Zentrum für Luft- und Raumfahrt e.V. entwickelt wurde. Zunächst werden in der Arbeit Grundlagen zum Thema Information Retrieval erörtert, sowie die wesentlichen verwendeten Technologien Apache Lucene und Eclipse Rich Client Platform beschrieben. Ausgehend von den Projektanforderungen beim DLR werden die Anforderungen an die Suchmaschine identifiziert und analysiert. Darauf eingehend wird der Entwurf des Systems dargestellt; im Anschluss daran wird die detaillierte Implementierung der entwickelten Volltextsuchmaschine beschrieben. Abschließend folgt eine Zusammenfassung der Arbeit sowie ein Ausblick auf weitere Tätigkeitsfelder.

This diploma thesis „Design and Development of a Search Engine for Scientific Documents“ describes the design and development of a search engine that has been implemented as part of a project at the German Aerospace Center (DLR). The first part of this thesis presents theoretical concepts of information retrieval (IR). Then, technological foundations are presented by describing the Apache Lucene library and the Eclipse Rich Client Platform. Starting with the project requirements from the DLR, a set of requirements for this work is identified and analyzed. Based on these requirements, the design of the system and the detailed implementation of the search engine is described. Finally, a summary of the work is presented and a prospect of potential extensions of the search engine is given.