

Aufgabenblatt 3: Indexterme, invertierte Listen

Aufgabe 1 Stoppwörter, Indexterme

Im folgenden sind eine Reihe von Titeln aufgeführt:

Speech retrieval in a multimedia system
Content-based retrieval in multimedia imaging
Photo-compact discs for multimedia systems
News on-demand for multimedia networks
Concept-based indexing and retrieval of multimedia documents
Wordspotting for voice editing and audio indexing
Imaging and multimedia: overview 1993
Supporting temporal multimedia operations in object-oriented database systems
Automated index generation for constructing large-scale conversational hypermedia systems
Filing and indexing: introduction
Pictorial information systems-prospects and problems
Bit group slice structure for signature file
Video indexing-an approach based on moving object and track
A flexible indexing structure supporting multi-attribute database application: MAIN
ADIIR: a testbed for experiments in document/image indexing and retrieval

- Unterstreichen Sie die Wörter, die Sie in einer Stoppwortliste vermuten würden.
- Bei welchen Repräsentationen ist die Verwendung von Stoppwortlisten sinnvoll und warum?
- Wie würden Sie die auftretenden Sonderzeichen behandeln? Warum?
- Welche Wörter würden Sie nicht in einem kontrollierten Vokabular zur intellektuellen Indexierung erwarten. Warum?

Aufgabe 2 Invertierte Listen

Konstruieren Sie mit einem Sortieralgorithmus eine invertierte Liste für mindestens die ersten drei Titel aus Aufgabe 1 und geben Sie die dazu erzeugten Listen an.

Aufgabe 3 Lemmatisierung

Führen Sie mit den Regeln von Kuhlen, wie sie in der Vorlesung behandelt wurden, an den folgenden Titeln eine Grundformenreduktion durch. Geben Sie die verwendeten Regeln an.

Photo-compact discs for multimedia systems
Concept-based indexing and retrieval of multimedia documents
Imaging and multimedia: overview 1993
Automated index generation for constructing large-scale conversational hypermedia systems
A flexible indexing structure supporting multi-attribute database application: MAIN
ADIIR: a testbed for experiments in document/image indexing and retrieval

Vergleichen Sie Ihre Ergebnisse mit denen des Finite-State Morphological Analyzers von Xerox (<http://www.rxc.xerox.com/research/mltt/Tools/morph.html>)

Aufgabe 4 Morphologische Analyse

Analysieren Sie den folgenden Satz morphologisch mit der in der Vorlesung angegebenen Methode:

Größere Dokumente mit Indexbäumen besser speichern.

Vergleichen Sie Ihr Ergebnis mit dem des Finite-State Morphological Analyzers von Xerox (<http://www.rsrc.xerox.com/research/mltt/Tools/morph.html>)