

Information Retrieval — Beispiel-Prüfungsfragen

Norbert Fuhr

- Einführende Fragen: Erzählen sie 3-5 Minuten etwas zu folgendem Thema. Gehen sie dabei auf die zu lösenden Probleme und die behandelten Lösungsansätze ein.
 - Wissensrepräsentation für Textinhalte
 - Evaluierung im IR
 - IR-Modelle
 - Prädikatenlogische Modelle
 - Ressourcen-(Datenbank-)Selektion
 - Visualisierung
 - Informationsextraktion
 - IR in XML-Dokumenten
- Wissensrepräsentation:
 - Welche Probleme der Freitextsuche werden durch welche Methoden beim informatischen Ansatz (/beim computerlinguistischen Ansatz) versucht zu lösen?
 - Welche Probleme der Freitextsuche werden durch die vorgestellten Verfahren nicht gelöst?
 - Freitextsuche vs. Dokumentations-sprachen
 - Welche der vorgestellten Verfahren erhöhen die Precision, welche den Recall?
- Evaluierung:
 - Welche Maße gibt es?
 - Wie kann man Recall, Precision und Fallout probabilistisch interpretieren?
 - Wie kann man den Recall abschätzen?
 - Wie entsteht eine Recall-Precision-Kurve?
 - Was muss man bei der Bewertung von Rangordnungen beachten?
 - Was ist die ‘Erwartete Suchlänge (expected search length)’, und wie kann man sie bei der Bewertung von Rangordnungen verwenden?
- Probabilistische Modelle:
 - Wie funktioniert Relevance Feedback?
 - Was heißt Optimalität beim PRP? Wie kann man das Ranking nach fallender Wahrscheinlichkeit rechtfertigen?
 - Geben Sie die Herleitung des BIR-Modells an!
 - Wie kann man die Parameter des BIR-Modells schätzen?
- Prädikatenlogische Modelle:
 - Beschreiben Sie die Grundkonstrukte von DAML+OIL
 - Wie kann man IR in Datalog modellieren (Hypertext, strukturierte Dokumente)?
 - Welche zusätzlichen Möglichkeiten bietet probabilistisches Datalog?

- IR-Systeme
 - Welche Ebenen kann man bei der Architektur von IR-Systemen unterscheiden?
 - Welche Stufen der Systembeteiligung gibt es?
 - Welche Arten von Suchaktivitäten kann man unterscheiden? Nennen Sie jeweils Beispiele!
- Implementierung von IR-Systemen:
 - Welche Zugriffspfade kann man für IR einsetzen?
 - Wie kann man effizient einen Text scannen?
 - Wie funktioniert boolesches/Vektorraum-Retrieval mit invertierten Listen?
 - Wie kann man invertierte Listen kompakt abspeichern?
 - Wie sind PAT-Trees aufgebaut, und welche Arten von Anfragen kann man damit prozessieren?
- Ressourcen-Selektion
 - Wie kann man optimale Datenbank-Selektion definieren?
 - Welche Arten von Kosten werden dabei berücksichtigt
 - Welche notwendigen Bedingungen ergeben sich aus der Definition des Optimums für den kontinuierlichen/den diskreten Fall?
 - Wie kann man bei (schwach) Kosten-monotonen Kostenfunktionen das Optimum berechnen?
- IR in XML-Dokumenten
 - Welche Bedeutung haben die verschiedenen XML-Standards für IR von XML-Dokumenten?
 - Welche Arten von Anfragen kann man bei XML-Retrieval unterscheiden?
 - Welche grundsätzlichen Methoden gibt es zu Retrieval für CO-Anfragen?