

Fragen	Punkte
1. Definieren sie Information Retrieval (IR) und grenzen sie IR anhand konkreter Beispiele von Datenbanktechnologie ab. Wie unterscheidet sich die Suche mit einem IR-System von der Formulierung einer Anfrage an ein relationales Datenbanksystem?	11
2. Beschreiben sie das prinzipielle Aufbauschema eines IR-Systems (Komponenten, <u>Arbeitsschritte</u>) Erläutern Sie dabei auch das Prinzip der invertierten Datei.	14
3. Erläutern Sie das Vektorraummodell für das Information Retrieval (Repräsentation, Retrievalfunktion). Wie unterscheidet es sich vom Booleschen Modell?	12
4. In einem auf dem Vektorraummodell basierenden IR-System liegen die Dokumentvektoren $D_1 - D_3$ sowie der Anfragevektor Q wie folgt vor: $D_1 = (0,8; 0,5; 0,3),$ $D_2 = (0,7; 0,6; 0,2),$ $D_3 = (0,6; 0,7; 0,6),$ $Q = (1; 0,3; 0,7).$ Berechnen Sie die Rangordnung der Ergebnisausgabe, wenn das Skalarprodukt als Retrievalfunktion verwendet wird und erläutern Sie dabei Ihre Vorgehensweise.	11
5. a) Wie unterscheidet sich Klassifikation von Clustering? b) Geben Sie ein praktisches Beispiel für den Einsatz von Klassifikationsverfahren im Text- und Dokumentretrieval.	10
6. Definieren und erläutern Sie das TFxIDF-Maß für die Dokumentindexierung. Welche Vorteile bietet es gegenüber der Indexierung nach dem Booleschen Modell?	10
7. Definieren Sie recall und precision als Evaluationsmaße. Beschreiben Sie, wie Sie den recall einer Suchmaschine wie Google praktisch bestimmen können.	10
8. Erläutern Sie das Prinzip des Google Pagerank-Algorithmus.	12
Summe	90

14