

Übungsblatt 11

28. Januar 2009

Hinweis: Soweit nicht anders angegeben, gibt es für jede korrekt bearbeitete Teilaufgabe einen Punkt. Die Abgabe der Hausübungen ist bis spätestens zum Beginn der nächsten Vorlesung möglich – entweder persönlich direkt vor der Vorlesung oder per Einwurf in den Briefkasten des Instituts (Informatikzentrum, zweiter Stock, vor Raum 238).

Aufgabe 25 (Mögliche Prüfungsfragen)

Zur Vorbereitung auf die Prüfung finden Sie unten typische Prüfungsfragen, die aus allen Themenbereichen der Vorlesung stammen. Beantworten Sie jede dieser Fragen kurz und prägnant (maximal fünf Sätze). Beschränken Sie sich also in Ihren Antworten auf die wesentlichen Aussagen (nehmen Sie sich Zeit zum Nachdenken!).

- a) Was haben das Boolesche Modell und das Coordination-Level-Match-Modell gemeinsam? Was ist der wesentliche Unterschied zwischen diesen beiden Modellen?
- b) Wie ist das TF-IDF-Maß definiert? Warum ist diese Definition sinnvoll?
- c) Was besagt das Probabilistic Ranking Principle?
- d) Im Gegensatz zur Precision sind Recall und Fallout schwieriger zu berechnen. Warum?
- e) Wie funktioniert Latent Semantic Indexing und warum hilft es beim Information Retrieval?
- f) Was besagt die Clusterhypothese? Halten Sie diese Hypothese für richtig?
- g) Was sind Language Models und was stellt man damit im Information Retrieval an?
- h) Was sind die wesentlichen Vor- und Nachteile von Pseudo-Relevance-Feedback?
- i) Was sind Support Vector Machines und wozu braucht man sie im Information Retrieval?
- j) Worin unterscheiden sich klassisches Information Retrieval und Web Retrieval?
- k) Was sind die wesentlichen Bestandteile einer Web-Suchmaschine und wozu dienen diese jeweils?