

188.412 VU 3.0 Information Retrieval 14.06.2007

1. Music-Retrieval: Beschreiben sie den Prozess zur Berechnung der Statistical Spectrum Descriptors sowie den Unterschied zu Rhythm Histograms betreffend der Art der Information die dargestellt wird. (5 Punkte)
2. Text-Indexing: Beschreiben sie unterschiedliche Methoden zur Feature Space Reduction (Pruning) bei hochdimensionalen Featurevektoren für überwachte und unüberwachte Textanalyse-Aufgaben. (10 Punkte)
3. ATC: Beschreiben sie die Qualitätsmaße Precision und Recall, sowie den Unterschied zwischen den Micro/Macro Formen derselben zur Evaluierung eines Classifiers, wie sie ermittelt werden, und was sie aussagen. (10 Punkte)
4. Information-Extraction: Was macht das Lexicon & Morphology Modul? Wann ist eher die Verwendung eines Lexicons, wann eines Morphology Moduls angebracht? (10 Punkte)
5. Text summarization: Was versteht man unter abstraktiver bzw. extraktiver Summarization? (10 Punkte)
6. QA: Nennen sie die gebräuchlichsten Komponenten eines Question Answering Systems. (5 Punkte)

188.412 VU 3.0 Information Retrieval 31.10.2007

1. Music-Retrieval: Beschreiben sie den Prozess zur Berechnung der Statistical Spectrum Descriptors sowie den Unterschied zu Rhythm Histograms betreffend der Art der Information die dargestellt wird.
2. Text-Indexing: Beschreiben Sie unterschiedliche Methoden/Arten von Feature Spaces die bei der Indizierung von Textdokumenten zum Einsatz kommen, wie sie aufgebaut werden, sowie ihre Eignung bzw. Vor/Nachteile. Beschreiben sie weiters die tfidf Methode zur Termgewichtung, wie diese Familie prinzipiell berechnet wird, und welche Annahmen dahinter stecken.
3. ATC: Beschreiben sie die Qualitätsmaße Precision und Recall, sowie den Unterschied zwischen den Micro/Macro Formen derselben zur Evaluierung eines Classifiers, wie sie ermittelt werden, und was sie aussagen.
4. Um Extraction Patterns zu definieren gibt es den atomaren und den Molekularen Ansatz. Beschreiben sie die beiden Ansätze. Welcher ist gebräuchlicher? Warum?
5. Was versteht man unter Coreferencing? Welche Arten von Coreferencing gibt es? Geben sie Beispiele an.
6. Information-Extraction: Was macht das Lexicon & Morphology Modul? Wann ist eher die Verwendung eines Lexicons, wann eines Morphology Moduls angebracht?