

Multimedia Information Extraction and Retrieval

Summer Term 2012

Exercise Sheet 2

Ralf Möller, Karsten Martiny

Exercise Session:
Thursday, April 26, 2012, 8.00-8.45, SBS95-D1025

1. Explain how a B-tree lexicon can be used to answer wildcard queries. What data structures are required to answer general wildcard queries (e.g. 'wil*ard')? How does the retrieval process work in this case?
2. Explain how data is stored in a permuterm index and how this structure supports the processing of wildcard queries.
3. Briefly explain the notions of *unstructured*, *semi-structured*, and *structured* data and give an example for each type of data.
4. Describe what *document fields* and *document zones* are and how these concepts can be used to enhance information retrieval.
5. Explain the notion of parametric search. What properties have to hold for a retrieval system in order to enable parametric search?
6. Explain how document zones can be used to score documents with respect to a certain query.
7. Explain the *bag-of-words model* in information retrieval.