

Midterm exam summary March 3, 2009

All material through basic index building today will be covered on exam.

Exam is open book with restrictions - see Announcements Web page

Exam out 4:20PM Tuesday, March 10 and due 3:00 PM Thursday, March 12, 2009

Bring questions Thursday (March 5)!

1

Major topics (highlights) to date

- Models
 - Modeling queries and documents
 - Satisfying, scoring and ranking w.r.t query
 - Vector models
 - LSI
 - Link analysis for ranking
- Evaluating retrieval systems
- Algorithms and data structures
 - Index structures
 - Evaluating queries using indexes
 - Building indexes

2

Reading covered on exam *Introduction to Information Retrieval*

- Chapter 1 – all sections
- Chapter 2 – all sections
- Chapter 3 – section 3.1: structures for dictionaries
 - Rest on wildcard queries and spelling correction
- Chapter 4 – Sections 4.1- 4.3
 - rest assigned but not on this test
- Chapter 6 - all sections except:
 - 6.1 zone scoring
 - 6.4.4 advanced dealing with document length in scoring

3

Continuation of Reading assigned to date in *Introduction to Information Retrieval*

- Chapter 7 – section 7.1 and 7.2 except:
 - 7.1.6 Cluster pruning
- Chapter 8 – all sections
- Chapter 18 – all sections
- Chapter 19 –all sections except:
 - 19.5 estimating relative sizes of search engine indexes
 - 19.6 finding near-duplicate pages
- Chapter 21 - all sections

4