

Summary for 2nd exam

DISTRIBUTED end of class Wed. April 29
DUE 5:30PM Friday, May 1

rules same as for first exam

All material **through** Monday April 20 will be covered on the exam.
Emphasis will be on **materials not covered on first exam**

1

Major topics (highlights) I emphasized in first exam

- **Models and methods**
 - Modeling queries and documents
 - Satisfying, scoring and ranking w.r.t query
 - Vector models for ranking
 - Link analysis for ranking
- **Evaluating retrieval systems**
- **Algorithms and data storage**
 - Index structures
 - Evaluating queries using indexes
 - Building “inverted indexes”
 - Storage concerns
 - Disk-based (external) algorithms
 - Compressing indexes

2

Major topics (highlights) II emphasized in **second exam**

- **Algorithms and data structures**
 - Index structures
 - Evaluating queries using indexes
 - Building indexes
 - **Storage concerns**
 - Disk-based (external) algorithms
 - **Compressing index structures** (at boundary 1st / 2nd exams)
 - Distributed computing for index construction and query evaluation

Emphasized
First
exam

3

Major topics (highlights) III emphasized in **second exam**

- **Discovering information**
 - Harvesting documents by Web crawling
 - Deep Web: finding information behind databases
- **Semi-structured information**
 - XML
- **Using user characteristics and feedback**
 - Search refinement
 - Personalization
 - Recommendation methods

4

Major topics (highlights) IV emphasized in **second exam**

- **Other analysis of information objects**
 - LSI
 - Clustering
 - Detecting near-duplicate documents
- **Non-text retrieval**
 - Images
- **Information from social networks**
 - Social network search
 - Aggregating social network data
 - Information from structure

5

Other end of term logistics

- **Project report due May 12**
 - You must submit a report
- **Project demonstrations**
 - **NOT** demonstrations for class
 - Meet with me, window **May 13 thru 18**
 - WASS calendar
 - **Discuss** project results and **clarify** points
 - **NO formal** presentation **unless have functionality to demo**

6