

COS 435

Summary for 2nd exam

All material **through** April 13 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**
 - Modeling queries and documents
 - Satisfying, scoring and ranking w.r.t query
 - Vector models
 - Link analysis for ranking
- **Evaluating retrieval systems**
- **Algorithms and data structures**
 - Index structures
 - Evaluating queries using indexes
 - Building indexes

2

Major topics (highlights) II

emphasized in [second exam](#)

- [Algorithms and data structures](#)
 - Index structures
 - Evaluating queries using indexes
 - Building indexes
 - Compressing index structures
 - Distributed computing for index construction and query evaluation
- [Semi-structured information](#)
 - XML

Emphasized
First
exam

3

Major topics (highlights) III

emphasized in [second exam](#)

- [Using user characteristics and feedback](#)
 - Search refinement
 - Personalization
 - Recommendation methods
- [Other analysis of information objects](#)
 - Clustering
 - LSI
 - Detecting near-duplicate documents

4

Major topics (highlights) IV

emphasized in [second exam](#)

- [Crawling the Web](#)
 - Harvesting documents by Web crawling
 - Characteristics of the changing Web
 - [Deep Web](#) Not on exam
 - [Non-text retrieval](#)
 - [images](#)
- 