COS 435
Summary for 2nd exam

All material through April 23 will be covered on the exam. Emphasis will be on materials not covered on first exam.

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

Major topics (highlights) II
emphasized in second exam

- Algorithms and data structures
  - Index structures
  - Evaluating queries using indexes
  - Building indexes
  - Compressing index structures
    - at boundary 1st/2nd exams
  - Distributed computing for index construction and query evaluation

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

Major topics (highlights) IV
emphasized in second exam

- Semi-structured information
  - XML
- Crawling the Web
  - Harvesting documents by Web crawling
  - Characteristics of the changing Web
    - Deep Web
    - Not on exam
- Non-text retrieval
  - Only basic ideas covered
    - images *