COS 435  2012 Midterm exam

DISTRIBUTED end of class Wednesday March 14.
DUE 3:00 PM Friday, March 16 in my office.
can be hand-written or typeset

All material thru compression is included (thru Mar. 5)

Exam is open book with restrictions:
• your personal notes and problem set submissions,
• the problem set solutions posted on the course Web site,
• Intro. to Information Retrieval by Manning, Raghavan, & Schütze,
• class notes and any of the readings posted on the “Schedule and Assignments” page of the Web site for the course

Outline of Major Topics Covered (highlights)

• 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