How Accurate is Item-Level Search for Digital Records?
Douglas W. Oard and Jason R. Baron

Digital Records
- Rapidly increasing scale
- Decentralized creation and management
- Somewhat “self-describing”
  - Content of individual items
  - Relationships between items

Key Questions
- How many “relevant” items exist?
- How many can practically be found?
- Which “methods” work best?
  - Which automated techniques?
  - Which ways of using those techniques?

Applications
- eDiscovery
- FOIA
- End-user search

The TREC Legal Track (annually since 2006)

~7 Million Scanned Business Records (Tobacco MSA)
- Title: CIGNA WELL-BEING NEWSLETTER - FUTURE STRATEGY
- Organization Authors: PMUSA, PHILIP MORRIS USA
- Person Authors: HALLE, L
- Document Date: 19970530
- Document Type: MEMO, MEMORANDUM
- Bates Number: 2078039376/9377
- Page Count: 2
- Collection: Philip Morris

~0.5 Million Corporate Emails w/attachments (Enron)

Two Types of Research Designs
- "Automatic" Task: Boolean, Ad Hoc, Relevance Feedback, Batch, Learning
- "Interactive" Task

Answer Keys for 116 eDiscovery “Production Requests”
All documents or communications that describe, discuss, refer to, report on, or relate to whether the Company had met, or could, would, or might meet its financial forecasts, models, projections, or plans at any time after January 1, 1999.