Next Generation Search

www.ir.iit.edu

Lab Director:
David Grossman
grossman@iit.edu
Past Accomplishments

- $4-6M in funding over last ten years
- Published two editions of “Information Retrieval: Algorithms and Heuristics”
- 2 Patents
- Several PhD students have graduated
  - Currently working at:
    - twitter.com
    - Lucent
    - BAE
- Key Sponsors
  - NSF
  - ARDA
  - AOL
  - Bit Systems
  - ASEC
Previous Efforts

• Query Log Analysis
• Web Search Evaluation
• Intranet Search Evaluation
• XML Search
• Scalable Document Clustering
• Arabic Name Search (ASOUNDEX)
An Interesting Result

• Last year, we were asked to test five major large-scale Boolean search engines.
• Average precision for a standard 500,000 document test collection was no higher than 0.22.
• In 1997 a group of research systems worked on this dataset. Average precision was roughly 0.25.
Effectiveness
2GB Collection
Default Stop Words / Stemming Enabled

Graph 4: 2GB Relevance Performance Topics 351-400
Current Work

- Name Search
- Extraction Improvement
- Complex Document Processing
Relevant Courses

- CS 422 - Data Mining
- CS 522 - Advanced Data Mining
- CS 429 - Information Retrieval
- CS 529 - Advanced Information Retrieval