



Directions for Search using Rich Language Processing

Justin Zobel

School of Computer Science and
Information Technology, RMIT University

G. Salton

- ◆ Automatic Index Organization and Retrieval, 1968

General purpose search



- ◆ Commercial systems work well and are thoroughly established.
- ◆ The gap between current performance and best possible performance is small.
- ◆ It is hard to demonstrate or identify significant improvements.
- ◆ Realistic data volumes are prohibitive for academia.
- ◆ Key data (clicks, queries, judgements) is not sufficiently available.

Opportunities?



- ◆ Question answering
- ◆ Fact extraction (query based)?
- ◆ Authorial support
- ◆ Querying of continuous data (such as transcripts)
- ◆ Speech data
- ◆ Summarization
- ◆ Reader-aware search
- ◆ Named-entity retrieval

Challenges



- ◆ Appropriate problem domains
 - Special-purpose collections
 - Specific applications
- ◆ Realistic data collections
- ◆ Theories that make testable predictions
- ◆ Well-defined evaluation procedures
- ◆ Well-defined problems
- ◆ Robust NLP (low-quality data)
- ◆ Scale, efficiency
- ◆ Shared tools