Evaluation in IR

Introduction to Information Retrieval
INF 141/ CS 121
Donald J. Patterson

Content adapted from Hinrich Schütze
http://www.informationretrieval.org

Outline

- Intro to Evaluation

- Standard Test Collections

- Evaluation of Unranked Retrieval

- Evaluation of Ranked Retrieval

- Assessing relevance

- Broader perspectives

- Result Snippets

Intro to Evaluation

- There are many implementation decisions to be made in an IR system

    - Crawler

        - Depth-first or breadth-first?

    - Indexer

        - Use zones?

        - Which zones?

        - Use stemming?

        - Use multi-word phrases? Which ones?

Intro to Evaluation

- There are many implementation decisions to be made in an IR system

  - Query

    - Ranked Results?

    - PageRank?

    - Which formula do we use in the TF-IDF Matrix?

    - Should we use Latent Semantic Indexing?

      - How many dimensions should we reduce?

Intro to Evaluation

- There are many implementation decisions to be made in an IR system

  - Results

    - How many do we show?

    - Do we show summaries?

    - Do we group them into categories?

    - Do we personalize the rankings?

    - Do we display results graphically?

Intro to Evaluation

- How can we evaluate whether we made good decisions or not?

    - Measure them

Measures for a search engine

- How fast does it index?

  - Number of documents per hour

  - Average document size

- How fast does it search

  - Latency as a function of index size

- Expressiveness of query language

  - Ability to express complex information needs

  - Speed on complex queries

Measures for a search engine

- We can measure all of these things:

  - We can quantify size and speed

  - We can make this precise

- What about user happiness?

  - What is this?

  - Speed of response/size of index are factors

  - But fast, useless answers won't make a user happy

- Need to quantify user happiness also.

Measuring user happiness

- Issue: Who is the user we are trying to make happy?

  - It depends.