Classic Information Retrieval

1

- Although search has changed, classic techniques still provide foundations – our starting point

2

Information Retrieval

- User wants information from a collection of "objects": information need
- User formulates need as a "query"
  - Language of information retrieval system
- System finds objects that "satisfy" query
- System presents objects to user in "useful form"
- User determines which objects from among those presented are relevant

3

Information Retrieval cont.

- Define each of the words in quotes
  - Information object
  - Query
  - Satisfying objects
  - Useful presentation
- Notion of *relevance* critical
  - What really want?
  - Insufficient structure for exact retrieval
- Develop algorithms for the search and retrieval tasks

4

Think first about text documents

- Early digital searches – digital card catalog:
  - subject classifications, keywords
- "Full text" : words + English structure
  - No "meta-structure"
- Classic study
  - Gerald Salton SMART project 1960's

5

Scaling

- What are attributes changing from 1960's to online searches of today?

  Some of answers we had on board:
  - Much much larger collections
  - Heterogeneous collections
  - Collections dynamic: docs come, go, change
  - Decentralized / distributed collections
  - More diverse users
    - Use for relevance?
  - More complex queries
  - Much much more computing resources
- How do they change problem?

6

Develop models

Begin with document models on board:

- Document is a _____ of terms*
  - Set
  - Bag
  - Sequence

* "term" is used instead of "word" to signal more general possibilities: serial numbers, nonsense, etc.

7