

---

## A Chequered History

- 1971: SMART System, Gerald Salton, Vector Space Model**
    - Information Retrieval (emph. on Relevance)
  
  - 1980s: Email, Electronic Junk**
    - Easier to rapidly generate information more than ever
    - Information Categorization
    - Information Routing
    - Information Filtering
  
  - 1991: Bellcore Workshop on Information Filtering**
    - Leads to Dec 1992 CACM Sp. Issue
    - Goldberg's article on Collaborative Filtering
  
  - [1992-95]**
    - Web Happens!
    - Maes: Social Information Filtering
  
  - 1996**
    - Berkeley Workshop on Collaborative Filtering
    - Leads to March 1997 CACM Sp. Issue on Recommender Systems
- 

---

## Shift in Models

- Information Retrieval**
    - [Terms, Documents]
  
  - Information Filtering**
    - [Features, Documents]
  
  - Collaborative Filtering**
    - [People, Documents]
  
  - Recommender Systems**
    - [People, Artifacts]
-

---

## Major Foci

- Information Retrieval (IR)**
    - Sparsity, Zipfian Distributions
    - Similarity by cosine measure
    - Precision, Recall, Tradeoffs, Linear Combinations
  - Immediate IR Concerns**
    - Synonymy, Polysemy
    - Phrases
    - Semantics and Roles
    - Computable Values
  - Relevance Feedback**
    - Query Expansion
  - “False hits are not necessarily bad”**
    - Preliminary HCI studies
  - Information Filtering**
    - query *becomes* profile, preference, feature
    - short-term operations *becomes* long-term
    - dynamic information needs *become* relatively static needs
    - selecting relevant info. *becomes* removing irrelevant information
- 

---

## Major Foci (Contd.)

- Some ideas from Dec 92 CACM**
    - LSI: Dimensionality Reduction [will encounter in Strand 1]
    - Collaborative filtering concept didn't really catch on immediately
  - Emphasis of Mar 97 CACM**
    - Information Filtering can be viewed as content-based filtering
    - Involving users can be beneficial (collaborative)
    - Still primarily online information (web pages, USENET articles, bookmarks)
  - Ways to Obtain Rating Information**
    - Explicit
    - Implicit: renting/buying actions, saving bookmarks, browsing, hyperlinks
  - Endemic Problems with Collaborative Filtering**
    - Clustering of Ratings, Preference Patterns, Diffusion
    - Sparsity
    - Effusivity, Negation
    - Day One (Cold Start)
    - Banana Problem
    - Preventing ratings mutiny
    - Scarcity of interaction in some domains (e.g. automobiles)
    - Free-riding
-