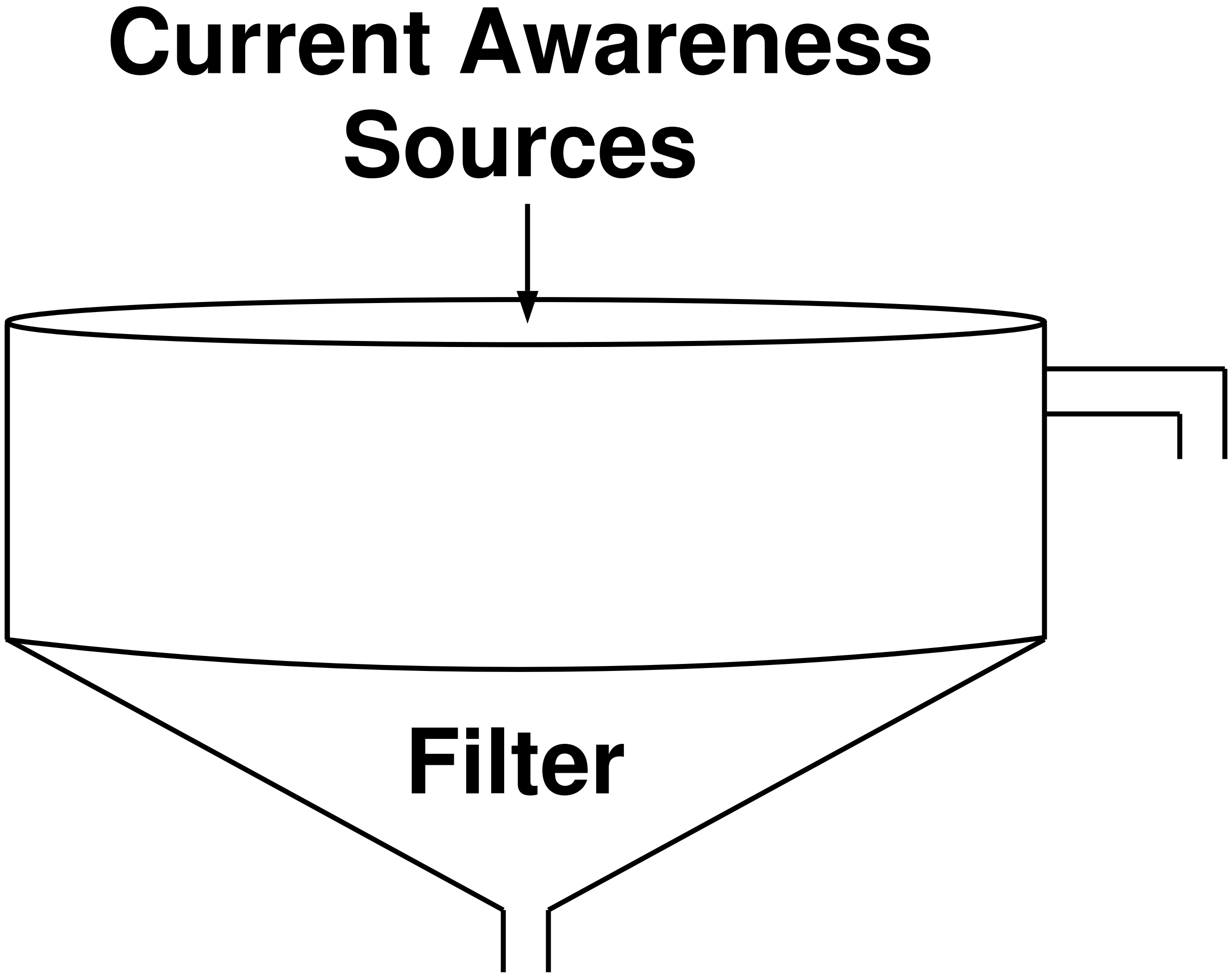
Information Access

		Question: old	Question: new
Information	old	Solved	Retrieval
Information	new	Filtering	Reality

Current Awareness + Filtering = Selective Dissemination of Information



Current Awareness

- **With Existing Filtering Systems**
  - Usenet News
  - Mailing Lists
  - Technical Reports
  - WWW Spiders
- **Emerging Applications**
  - Electronic Journals
  - Digital Library Accessions
  - MBONE Archives

Basis for Selection

- **Content**
  - Words and Phrases
  - Structure
- **Evidence**
  - Rules or Programs
  - Explicit User Judgements
  - User Behavior
- **Collaboration**
  - Shared User Judgements

Text Retrieval

- **Boolean Model**
  - Rule-Based
- **Vector Space Model**
  - Natural Language
- **Probabilistic Model**
- **Connectionist Networks**
  - Machine Learning

A Proposed Taxonomy

- **Filtering Approach**
  - Content-Based
  - Collaborative
- **Architecture**
  - Distributed
  - Centralized
- **Knowledge Acquisition**
  - Rule-Based
  - Machine Learning
    - Explicit Feedback
    - User Observation

Filtering Systems

http://www.glue.umd.edu/enee/medlab/filter/filter.html

System	Domain	Avail	Basis	Acquisition	Architecture
SIFT	News and TRs	Yes	Content	Natural Language	Centralized
Procmail	Mail	Yes	Content	Rules	Distributed
Lurker	News	Yes	Content	Rules	Distributed
GHOSTS	News	Unknown	Content	Rules	Distributed
NewsClip	News	Yes	Content	Program	Distributed
Browse	News	Yes	Content	Explicit Feedback	Distributed
NewsWeeder	News	Limited	Content	Explicit Feedback	Distributed
InfoScope	News	No	Content	User Observation	Distributed
MAXIMS	Mail	Yes	Collaboration	User Observation	Distributed
Maltz RN	News	Unknown	Collaboration	Explicit Feedback	Centralized
WebHound	WWW	Yes	Collaboration	Explicit Feedback	Either
URN	News	Unknown	Both	Explicit Feedback	Both
Tapestry	Mail and News	No	Both	Rules	Both

Performance Evaluation

- **Effectiveness**
  - Precision
    - Relevant Selected / Total Selected
  - Recall
    - Relevant Selected / Relevant Available
  - User Control
  - Learning Rate
- **Efficiency**
  - Computational Complexity
  - Network Bandwidth
  - Storage

Research Directions

- **Learn from human filtering behavior**
  - Collaboration seeds content search
  - Content search informs collaboration
- **Exploit multiple sources of evidence**
  - Unintrusive observation
  - Explicit feedback
  - Direct manipulation
- **Develop distributed architectures**
  - Required for large-scale applications