Documents are networked into information spaces
- World Wide Web
- Blog space
- Scientific and Medical Literature
- Patents
- Common Law

2. What can we do with documents now?
- Crawl the network to collect documents
- Store and access documents
- Search them by keyword
- Citation indexing (who cites me?)

3. What else is feasible?
- Extract keywords and terms (Natural Language Processing)
- Linguistic rules
- Frequency of occurrence
- Summarize documents & collections (Machine Learning)
- Extract important keywords
- Extract important sentences

4. Open Problems
- Extract Metadata (Semantic Web)
- Disseminate documents over Peer-to-Peer-networks
- Extract topics, concepts, taxonomies
- Organize Desktop document collections
- Organize Group document collections

5. Desktop of the Future
- The computer becomes:
  - Personal library (books, articles, news, ...)
  - Personal filing space (finance, receipts, ...)
- Material automatically organized, to be:
  - Searchable
  - Browsable

6. Peer-to-Peer Document Management
- Members of a social network share their documents
- Profile of each member, based on his/her collection, formed automatically
- Members connected in a peer-to-peer network
- Documents exchanged automatically based on member profiles

7. Peer-to-Peer Document Management
- Members of a social network share their documents
- Profile of each member, based on his/her collection, formed automatically
- Members connected in a peer-to-peer network
- Documents exchanged automatically based on member profiles

8. Peer-to-Peer Document Management
- Members of a social network share their documents
- Profile of each member, based on his/her collection, formed automatically
- Members connected in a peer-to-peer network
- Documents exchanged automatically based on member profiles

9. Automatic Topic Extraction
- Given a document collection
- Find the key topics
- Describe topics by sets of keywords
- Documents can have more than one topic
- Achieved by unsupervised learning of word-topic associations

10. Automatic Topic Extraction
- Given a document collection
- Find the key topics
- Describe topics by sets of keywords
- Documents can have more than one topic
- Achieved by unsupervised learning of word-topic associations

11. Experience Management
- Capture experience of systems administrator
- Collect previous problem cases and solutions
- Given a new problem and its symptoms
- Retrieve closely related solved cases (automatically)
- Modify related cases to solve new problem (manually)

12. Where are we heading?
- Information access is becoming free.
- We are building tools to deal with information overload.
  - Information retrieval
  - Text mining
  - Link mining
  - Text collection organization

Machine Learning and Networked Information Spaces
http://www.cs.dal.ca/~eem/malnis