ClusterBook, a Tool for Dual Information Access

Gheorghe Mureșan, David J. Harper, Ayşe Göker, Peter Lowit
School of Computer and Mathematical Sciences
The Robert Gordon University
Aberdeen, UK
{gm, djh, asga, pl}@scms.rgu.ac.uk
http://www.scms.rgu.ac.uk/research/ir

The work is part of the WebCluster project, sponsored by Ubilab, Zurich.
The library metaphor

1. Select library
2. Consult catalog
3. Browse shelves
4. Use inter-library scheme

Information Need Formulation

Information need
1. Select source collection

2. Explore source collection with ClusterBook

3. Search WWW

The model of the search process
The client-server architecture

Client (ClusterBook)

Query:

1. GetCollections
2. GetClustering
3. SearchClusters
4. SearchDocuments
5. GetClusteredHits
6. ...
7. ...

Mediated query:

Server

GetHits

Informia meta-search engine

1. Indexer
2. Clustering Framework
3. Cluster-based searcher
4. Ranked-based searcher

1. Document collection
2. Vocabulary, index, inverted file
3. Cluster hierarchy

Client (ClusterBook) Server

Informia meta-search engine
Critical issue: The label generation

1. Document representatives
   • searching

2. Cluster representatives
   • browsing
   • searching
   • mediation

3. Collection representatives
   • collection selection

Fixed Plants

- Coastal Wind Farms
  - Pacific Rim Wind Farms
  - Design of Coastal Wind Farms
- Inland Wind Farms
  - Desert Wind Farms
  - Design of....

Power Generation

- Wind Energy
- Propulsion

Portable Generators

- Wind generators for yachts
Contributions of ClusterBook

- Dual information access to a document collection via:
  - Hierarchic structure based on document-document similarity
  - Linear view based on query-document similarity

- Support for a variety of search tasks:
  - Following an information-gathering plan
  - Exploring a topic in an undirected fashion
  - Monitoring a topic or aspects of a topic

Contributions of WebCluster

- Propose and explore system-based mediated access to very large heterogeneous document collections

- Explore the use of language models for building cluster and document representatives

- Offer a framework for building structured portals on the WWW

- Offer a framework for building collaborative environments
Moody's Investors Service Inc said it lowered the debt and preferred stock ratings of USX Corp and its units. About seven billion dollars of securities is affected. Moody's said Marathon Oil Co's recent establishment of up to one billion dollars in production payment facilities on its prolific Yates Field has significant negative implications for USX's unsecured creditors. The company appears to have positioned its steel segment for a return to profitability by late 1997, Moody's added.

Ratings lowered include those on USX's senior debt to BAA-1 from BAA-3. Reuters &#39;