Clustering Metadata for improved Querying

Patrick Trinkle, UMBC
tri1@umbc.edu

24” x 16”

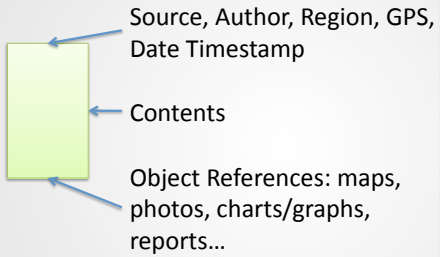
Goals

1. Glean information from GUI rather than searching through endless lists.
2. Change the way people who read reports find collaborative and augmenting information.
3. Facilitate discovery of new information.

Input/Process

Corpus

•Reports about one or several closely related subjects.
•Reports contain metadata.

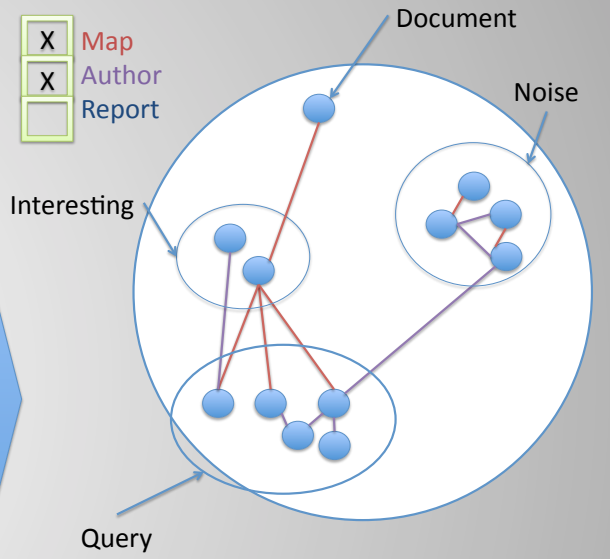


Process

1. Extract Metadata and keyword information.
2. Build vectors with metadata and references.
3. Group documents by features; documents with more features/content in common are in small clusters.

*Clustering algorithm can be chosen based on experimental performance.

User Interface



•Demo Graph is 2-dimensional, people can interact in up to 3 Euclidian dimensions.
•Distance between points indicates how closely they’re related.
•The style of the lines between documents specify type of relationship.

Previous Work

•Scatter/Gather – list tree view, hierarchy.
•Cluster Interface for scientific data.
•Metadata Extraction – automatically grab features.
•Normal Listing Engines – lists of results.
•Co-occurrence – reframing the problem.

Future Work

•Build system.
•Effectiveness with different data sets?
•Other problems addressable with system?