

Abstract

Information exploration involves building bridges between *document space* (the realm of information artifacts, ranging from books to e-mail to bibliographic citations) and *concept space* (the context and knowledge required to interpret documents). This dissertation proposes an interactive, structure-mediated approach to exploring information in a heterogeneous, distributed environment—such as a digital library. It presents a conceptual model for information exploration, a system (SenseMaker) that embodies the proposed model, and several evaluative user studies.

Two principles form the crux of our structure-mediated conceptual model, which supports users in moving back and forth between document space and concept space:

- Users should be able to build document-based structures for document collections by interactively organizing them around contextual dimensions (e.g., author). These higher-level structures reveal key concepts (e.g., specific authors).
- Users should be able to create new collections of documents by sculpting or adding to document-based structures. This reuse of structure enables concept-based filtering and concept-based finding of documents.

This dissertation introduces *iterative bundling* and *flexible duplicate detection* as two strategies for interactively building document-based structure. *Iterative bundling* creates aggregates of similar documents. It allows the definition of similarity to vary according to contextual dimension. For example, a different metric might be used for bundling by author than for bundling by topic. *Flexible duplicate detection* identifies equivalent documents and merges them together. Again, the definition of equivalence can vary.

The creation of a document collection is a prerequisite to building document-based structure for a collection. This dissertation defines an ontology of collection-creating actions that includes *specifying* the characteristics of a new collection, *expanding* a previous

collection, *limiting* a previous collection, and *returning* to a previous collection. Collection-creating actions that utilize structure fit into several of these categories.

The SenseMaker system embodies the proposed structure-mediated conceptual model. SenseMaker utilizes protocols and other technologies developed for the Stanford Digital Library project. Its interface design introduces the concept of *hi-cites*, a user-interface construct to facilitate browsing.