

Using Superimposed and Context Information to Find and Re-find Sub-documents

Sudarshan Murthy¹, Uma Murthy², Ed Fox²

¹Portland State University, ²Virginia Tech.

What?

- Explore the use of superimposed information and context information to **find and re-find sub-documents**
- **Superimposed information** is new information such as annotations placed over existing information
- **Context** is information related to a sub-document, e.g., text surrounding a search term

Why?

- Considerable user effort to find relevant sub-documents with current tools; finding effort is repeated when re-finding
- Search engines return documents, but users frequently need sub-documents: Users routinely perform **intra-document searches**, e.g., search within a Microsoft® Word document, to find sub-documents
- Some search engines return snippets, but they could do better

How?

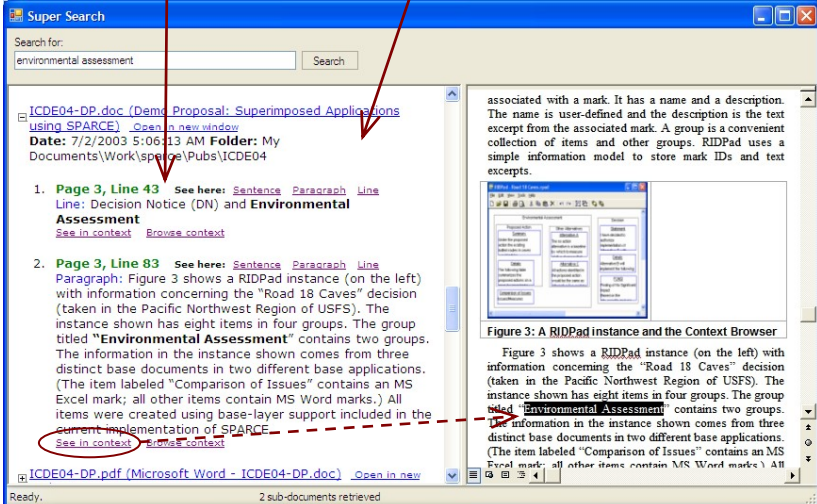
- **Return documents and sub-documents**, provide richer context information
- Unify document and intra-document search
- **Support navigation directly to sub-documents** from search results
- Leverage finding effort when re-finding information

Results so far:

- **SPARCE**: Infrastructure to reference sub-documents and retrieve context information, and a few superimposed applications using SPARCE
- **Super Search** (*Sub Search*) prototype: Augments documents returned from a search engine with references to sub-documents and adds context information; factors in previously found information

Selected context information for sub-documents

Documents and sub-documents returned for a query. The sub-document list is numbered.



associated with a mark. It has a name and a description. The name is user-defined and the description is the text excerpt from the associated mark. A group is a convenient collection of items and other groups. RIDPad uses a simple information model to store mark IDs and text excerpts.

Figure 3: A RIDPad instance and the Context Browser

Figure 3 shows a RIDPad instance (on the left) with information concerning the "Road 18 Caves" decision (taken in the Pacific Northwest Region of USFS). The instance shown has eight items in four groups. The group titled "Environmental Assessment" contains two groups. The information in the instance shown comes from three distinct base documents in two different base applications. (The item labeled "Comparison of Issues" contains an MS Excel mark; all other items contain MS Word marks.) All items were created using base-layer support included in the current implementation of SPARCE.

Figure 3 shows a RIDPad instance (on the left) with information concerning the "Road 18 Caves" decision (taken in the Pacific Northwest Region of USFS). The instance shown has eight items in four groups. The group titled "Environmental Assessment" contains two groups. The information in the instance shown comes from three distinct base documents in two different base applications. (The item labeled "Comparison of Issues" contains an MS Excel mark; all other items contain MS Word marks.) All items were created using base-layer support included in the current implementation of SPARCE.

Go directly to a sub-document