

---

# **Executive Summary**

*Excerpted from*

## **The Value of CSA Deep Indexing for Researchers**

Dr. Carol Tenopir  
Dr. Robert J. Sandusky  
School of Information Sciences  
University of Tennessee, Knoxville

Margaret M. Casado  
John C. Hodges Library  
University of Tennessee, Knoxville

October 30, 2006

---

### **Introduction**

In 2005-2006 CSA developed a new “deep indexing” method to extract tables and figures from journal articles, index each table and figure, provide a retrieval method to locate tables and figures or complete articles that contain relevant figures or tables, and link them back to the article. A prototype Tables & Figures index of approximately 300,000 objects was made available for testing on the CSA Illumina Web-based platform in April 2006. The Tables & Figure index prototype was particularly strong in its coverage of biological, environmental, and aquatic sciences.

Each table and figure extracted from a journal article is assigned index terms as appropriate for the type of table or figure (photograph, histogram, map, etc.), subject indexing, geographic indexing, taxonomic indexing, statistical indexing, and other relevant data using an automated indexing system. All tables and figures in an article are fully indexed and can be searched separately. A combined “enhanced abstract” display shows all of the tables and figures within a single article along with bibliographic, full indexing, and abstract for the article. These enhanced abstract articles can be linked to the electronic journals collection of a library. Tables and figures are displayed within the CSA deep indexing system as thumbnails, most of which can be displayed in an enlarged format.

CSA asked a research team at the University of Tennessee, Center for Information Studies, to test the utility of deep indexing for scientists and explore how it might enhance scientific research. The team was led by Professor Carol Tenopir, with Donald W. King (both wellknown information industry researchers and co-authors of *Towards Electronic Journals*), Dr. Robert Sandusky (participant in digital library research on disaggregation of journal articles), Margaret Casado (instructional librarian), and CIS staff and graduate students.

## **Methods**

CSA identified librarians at universities and research institutes in Europe and North America who would assist with the recruitment of scientists to test the system. In all, sixty scientists in 9 organizations participated (7 universities and 2 research institutes; 3 in Europe and 6 in the United States).

One member of the research team visited each of the participating organizations, to provide introductory sessions, gather data, distribute passwords, and provide instructions on additional data collection. Multiple methods of data collection allow data validation and triangulation for both quantitative and qualitative data. They allow the team to study both predictive questions, such as how might indexing of tables and figures might be used by scientists, and functional questions, such as what types of search and interface features are particularly useful for a tables and figures system.

Data collection methods included: pre- and post-search questionnaires to describe potential usefulness, expectations, and current practices; observation sessions to discover, through initial and real-time interactions with the system, potential usability and functionality issues; and structured diaries of searches performed by the participants, on topics of their own choosing, in the weeks following the introductory sessions to gather more detail on potential uses of the Tables and Figures index prototype, encourage additional participant experiences with the system, and identify both useful functions and concerns with the prototype.

## **Conclusions**

Questionnaires reveal that these participants are frequent readers of electronic journal articles. They use multiple systems to locate relevant articles, among which, electronic indexing and abstracting services and web search engines are used the most. Most rate access to databases and relevance of items retrieved as essential attributes of online search services, with ease of searching rated next in importance. Ratings of satisfaction with these attributes were slightly lower than ratings of importance, showing that there is room for improvement in the systems and services they use. Lack of access to high quality versions of full text articles of any publication date – in the past as well as the most recent articles – are the most common complaints. Only a few currently use systems for images, most often web search

engines and most often for maps or photographs, although a majority rated the importance of access to high quality graphs, figures, or tables as essential for both teaching and research.

After using the Tables & Figures index prototype, participants rated access to relevant articles as the most important attribute of the system, although satisfaction with this attribute is rated lower, probably due to the small size and scope of the prototype. On average the rating of the importance of various types of objects for teaching and research went up slightly after exposure to the prototype system.

A total of 378 searches were recorded in the diaries of the 60 participating scientists, and many different types of searches were performed. Scientists were most often “exploring a topic” or “seeking specific information within a topic,” followed by seeking a specific author and seeking a specific item or article. Most felt the objects were most useful within the context of the entire article and some even cautioned about using a table or figure out of context. Complete captions, a high quality reproduction of the table or figure, and a link to the full text are important in most cases, although photographs and maps are more likely to be able to stand on their own.

Diary entries revealed that participants believed that in a majority of these searches (60%) locating the same information would have taken longer in a traditional database, although in most cases (63%) they believed they would have found the information eventually in traditional searches. Even though they were searching a prototype database with limited coverage, nearly two-thirds of the searches located relevant information.

Scientists identified many potential uses of tables and figures indexing to their work in both the observational sessions and diary entries. These potential uses include:

- teaching/lectures/presentations for which they would download figures directly into presentation software
  - locating and retrieving data in particular formats or particular object types
  - making comparisons between their work and the work of others
  - gaining faster and more precise understanding of the work reported in articles by direct examination of the tables and figures
  - assistance with writing of review papers, meta-analysis, proposals, and generating hypotheses
  - improving the efficiency of searching by providing more precise and smaller results sets
  - supporting the transformation of practice and supporting the learning of new skills and methods, including how to effectively present results in tables, figures, and graphs
  - use of objects by librarians to directly answer reference questions
- Overwhelmingly, respondents said that the ability to search for specific types of objects would make a difference in their search and discovery processes. They believe it will save time, allow them to work more efficiently, aid in presentations, and allow them to find more relevant results.

Almost all participants (96%) prefer a system that combines the features of Tables and Figures indexing with traditional searching. They ranked the importance of specific features, such as the ability to limit searches to particular object types (e.g., figures and tables) as highly important; they also gave the highest satisfaction ratings to the implementation of search limits for figures and tables.

Overall, the scientists in this study found many potential uses for a system that indexes tables and figures. It can be used to help them find information and specific objects for teaching, presentations, research, and learning. The utility of such a system is highest if the functions are incorporated into a traditional system, linked seamlessly to full text articles, and include high quality images with complete captions.

#