University of Tennessee, Knoxville

From the SelectedWorks of Carol Tenopir

2007

Uses of figures and tables from scholarly journal articles in teaching and research

Carol Tenopir, University of Tennessee - Knoxville
Robert J Sandusky, University of Tennessee - Knoxville
Margaret M Casado, University of Tennessee - Knoxville

Available at: https://works.bepress.com/carol_tenopir/37/
Uses of Figures and Tables from Scholarly Journal Articles in Teaching and Research

Robert J. Sandusky (corresponding author)
School of Information Sciences, University of Tennessee, 451 Communications Building, 1345 Circle Park Drive, Knoxville, TN 37996 Tel: +1 865-974-2785 sandusky@utk.edu

Carol Tenopir
School of Information Sciences, University of Tennessee, 451 Communications Building, 1345 Circle Park Drive, Knoxville, TN 37996

Margaret M. Casado
John C. Hodges Library, University of Tennessee, 135 John C Hodges Library, 1015 Volunteer Boulevard, Knoxville, TN 37996

This paper describes how scientists utilize specific journal article components, the tables, figures, maps, photographs, and graphs contained in journal articles, to support both their teaching and research. These findings are taken from a comprehensive investigation into scientists’ satisfaction with and use of a prototype retrieval system that indexes tables and figures culled from scientific journal articles. Rather than focusing on seeking and searching, this paper summarizes four ways in which scientists utilize the information they find in tables and figures obtained from journal articles. While the first type of use described here, creating new fixed documents, confirms the findings of previous research, the other three types of use reveal emerging practices with journal article components: creating documents to support performative activities; making comparisons between a scientist’s own work and the work of other researchers; and creating other information forms and objects.

Introduction

Most previous research on scientists’ use of journal articles has focused on how scientists seek, search for, and select entire articles to support research-related activities, such as maintaining current awareness, writing papers and grant proposals, and investigating new research areas (Friedlander, 2002; Institute for the Future, 2002; Tenopir & King, 2004; National Science Board, 2006). For the most part, those studies give little or no attention to the details of how scientists use those articles and their components in their research activities. A smaller number of published reports have commented upon how scientists use journal article components, such as tables, figures, and reference lists, but primarily with respect to making relevance assessments for entire articles and almost exclusively in the context of scientists’ research activities (Stewart, 1996; Bishop, 1999; Stelmaszewskak & Blandford, 2004). We begin to see a more complete picture of the ecology of document and component use by taking a wider view of scientists’
interactions with documents and specific journal article components in multiple work contexts, including teaching and research. Development of a more complete map of scientists’ interaction with and use of documents and components can help systems and service providers design appropriate technology (O’Hara et al, 1998).

This paper discusses many of the uses scientists make of the tables, figures, maps, photographs, and graphs contained in journal articles in both teaching and research. These findings are taken from a comprehensive investigation into scientists’ satisfaction with a prototype retrieval system that indexes tables and figures culled from scientific journal articles (Sandusky & Tenopir, forthcoming). This paper summarizes four ways in which scientists utilize the information they find in tables and figures obtained from journal articles: creating new fixed documents; creating documents to support performative activities; making comparisons between their own work and the work of other researchers; and creating other information aggregations, compilations, forms, and objects.

Few systems have provided direct access to tables, figures, or other journal article components in a manner similar to the journal article component indexing prototype evaluated in this study. The DeLIver testbed, in operation circa 1996-1998, provided end-users with the capability to retrieve articles by indexing information contained in specific components within complete documents. DeLIver supported then-innovative search capabilities including searching of section headings, table and figure captions, table text, and cited references (Bishop, 1999; Bishop et al, 2000). More recently, Lu et al. (2006) developed a system for categorizing figures automatically extracted from PDF documents sampled from the CiteSeer digital library. Their system uses a machine learning approach combined with automatic image processing algorithms to support the classification of extracted figures into five categories: photograph, 2-D plot, 3-D plot, diagram and others. While Lu et al. compared the performance of their figure classifier to human classification of the same test set of more than 5,000 figures extracted from randomly selected scientific papers from CiteSeer, no user-based evaluation of the efficacy of their approach has been attempted. Liu et al. (2006), in a related effort, describe a system that automatically extracts table metadata from PDF documents in CiteSeer by converting PDF documents to formatted text files, detecting table “candidates,” and identifying table data (e.g., frame, metadata, layout, content metadata).

Investigations of scientists’ use of journal articles for purposes other than research have been rare. Hart (1998) surveyed faculty at one non-research intensive university to investigate the relationship between faculty roles of teaching, research, and service and six abstract categories of information sources, but did not investigate how faculty searched for specific forms of information; nor did he examine how they made use of the information they found. Borgman et al (2005) examined the general information seeking behavior of geography faculty in support of undergraduate teaching as part
of the ongoing evaluation of the ADEPT (Alexandria Digital Earth Prototype) project. Borgman and her colleagues, noting linkages between information seeking for teaching and research, found that geographers “(a) ...found it easier to articulate their research activities than their teaching activities; (b) research and teaching activities were viewed as mutually reinforcing; (c) they continually scan their environment for information sources and glean for both purposes; and (d) they rely on their own research data as information sources for teaching.”

The journal article component prototype evaluated here is designed to give end users direct access to disaggregated journal article components. The prototype provides an end-user search interface supporting construction of complex Boolean queries and selection of a variety of field limits and results filters. For example, the searcher can enter search terms and limit the string matching to figure or table caption text, or geographic, statistical, or taxonomic descriptors. The user can likewise limit the results returned to figures, tables, graphs, maps, and/or photographs. Figures 1 through 3 show portions of the prototype interface, including the results set display (Figure 1), a portion of the surrogate record corresponding to one figure (Figure 2), and the journal article’s “enhanced abstract,” which provides hyperlinked thumbnails of each of the figures and tables extracted from that article (Figure 3).
Figure 1: Display of two items in a results set returned by the component prototype developed by ProQuest CSA. Each item in the results set includes a truncated caption, a thumbnail of the figure or table, citation information (author, journal name, volume, issue, page numbers), and links to the complete component surrogate (see Figure 2), the containing article’s enhanced abstract (see Figure 3), a link to the article full text, and the descriptors assigned to the table or figure’s surrogate record. (Note: At the time the component prototype was evaluated, the article title was not displayed as part of the results set due to a design oversight.)

Figure 2: Display of a portion of the component’s complete surrogate record in the component prototype developed by ProQuest CSA. The figure is classified as a schematic. Note that along with a higher resolution image of the component and the full caption, the surrogate displays general (Atrioventricular sulcus, etc.), geographic (Norway), and taxonomic (Salmo salar) descriptors.

Figure 3: Display of the component prototype’s enhanced abstract for the entire article. Note the addition of hyperlinked thumbnail images to four figures and one table at the bottom of Figure 3. Selection of one of the thumbnails causes the display of the complete surrogate for that specific table or figure (as illustrated in Figure 2).

Previous research has shown that figures and tables are important journal article components because of their ability to succinctly summarize the article content, including data, methods, and results. The current study advances our understanding of how disaggregated journal article components are utilized by scientists in both research and teaching.

Methods

Multiple methods were employed to collect both qualitative and quantitative data for this study. Data collection procedures and instruments were designed to provide insights into the potential impacts of the indexing and retrieval of journal article components on teaching and research. One of three researchers traveled to each of the nine research sites to introduce the system and its capabilities, introduce the research plan, and conduct observations of scientists’ use of the journal article component indexing prototype. The on-site sessions included distribution and collection of the pre-search questionnaire (to measure characteristics of the participants, prior knowledge, experiences with, and potential uses of journal article component indexing), an introduction to the journal article component indexing approach and the prototype, distribution of the structured diaries, and finalization of times and locations for the individual observations. The three research team members conducting the site visits used the same presentation materials in order to minimize the differences in how the participants were introduced to the project.

There are some obvious disadvantages to providing the participants with an introduction to and hands-on practice with the system features under evaluation as bias is introduced. It may be expected that the participants would have a heightened awareness of the new system features compared to their awareness of the same features in a more natural research setting. However, given the time constraints and the limited nature of the research questions posed, this tradeoff was deemed acceptable. This study is not designed to be a situated, naturalistic study of the use of this system in the life of scientists; nor is designed to be a comprehensive study of the information practices of scientists in a wider context.

During the introductory onsite session, participants were asked to volunteer to return for a one-on-one observation session and most agreed to do so. The observations were conducted based upon the recommendations of Monk et al (1993) for conducting an “obtrusive” observation of user behavior. Participants are provided with a short set of “typical” system tasks plus some open-ended tasks and asked to “think aloud” as they use the system, followed by a short debriefing in order to ascertain participants’ impressions of the journal article component prototype, how its features might or might not benefit their work in the future, detect usability problems, and allow participants to offer suggestions for improvement.
Structured diaries (Hyldegård, 2006) were used to probe more deeply into how journal article component indexing might be useful to researchers. Each participant was asked to conduct a minimum of five and up to ten searches on the prototype during the two weeks following the site visit using a unique login ID and password. They could search on any topic of interest to them, using any of the features available on the system. The diaries mixed structured questions and open-ended questions, and sought to capture users’ queries and query modifications and the outcomes of their searches.

At the end of the study, after participants turned in their structured diaries, a post-search questionnaire was distributed electronically. The post-search questionnaire contained some of the same questions as the pre-search questionnaire to examine how exposure to the journal article component prototype influenced participant’s perceptions of the utility of journal article component indexing in their work.

Using multiple methods increases the validity of the measurements because data about the research questions is acquired and interpreted from complementary perspectives. Reliability, or the likelihood that the questions we are asking will be accurately and consistently answered by respondents, is increased by measuring the same things in more than one way. Reliability is also improved by using questions about journal reading tested and asked many times in prior research by Tenopir & King. Individually, each method has particular advantages and disadvantages and thoughtful combination of methods allows us to address limitations of one method with data gathered using a complementary method. This allows researchers to triangulate, or draw inferences or conclusions about the phenomena being studied, with more confidence than relying on a single method (Bryman, 2003).

Both the reliability and validity are hampered by the fact that participants interacted only with a prototype system. The size of the database was limited, some of the tables were illegible, and many table / figure captions were truncated. The results must be interpreted with these limitations in mind. Some participants answered questions based on what they believed the prototype system could become, while others were distracted by the limitations of the prototype and answered questions based on those limitations, which are not likely to be a part of an operational system. While the relatively small sample and its nature as a sample of convenience limit the ability to generalize from the data reported here, the results are important in revealing the extent of use of tables and figures drawn from scholarly journal articles.

ProQuest CSA, the sponsor of this project, contacted institutions that would likely provide access to participants for this study. ProQuest CSA’s contacts at each institution then recruited individual researchers at their institution. The institutions selected were a mix of universities and research institutes located in the United States (5 universities and 1 research institute) and Europe (2 universities and 1 research institute). Each institution recruited

between four and twelve scientists and researchers, for a total of sixty participants. This sample of convenience yielded a group of participants representing a cross section of science subject disciplines, geographic spread, and academic level. Although it is not random, the number of institutions and participants at each institution represents an adequate sample from which to draw meaningful, if provisional, conclusions. Participants self-identified by providing their academic rank and/or job title to provide a rough indication of each participant’s research experience. A plurality of participants identified themselves as either professors or researchers, denoting a high level of research experience. The participant pool also included a large number of post-doctoral researchers. The next largest group was “students,” who all held at least a bachelor’s degree. Librarians, while not the primary focus of this project, represent an important constituency because they act as intermediaries and are often responsible for the acquisition, promotion, and training of users for all kinds of searching and abstracting and indexing systems at their institutions.2

Participants also provided information on their highest-level academic degree. Two-thirds hold a PhD or MD. Masters degrees are held by nine participants, bachelors degrees by nine participants, and two participants did not respond to this question. We asked participants to provide the disciplinary label with which they most closely identify. Various sub-fields of biology were grouped together as “Biology” and comprise a near-majority. The discipline occurring with the second-highest frequency is “Ecology”. All others occurred at a frequency of five or fewer. In summary, most participants hold a PhD or similar terminal degree; most are working in the life sciences; all are serious consumers of research and users of electronic search systems; and most are also producers of primary scientific research and journal articles. Research consumes the majority of the time of three-quarters of the participants, and over half of the participants spend up to half of their time on teaching.

Results

Journal article component use takes many forms, four of which are discussed here. While one of these, the creation of new fixed text documents, is expected, the other three uses discussed here have been infrequently or never before reported in the LIS literature.

Disaggregated components can be used to support the creation of new, fixed text documents. Examples of fixed text documents are journal or conference papers; research proposals; meta analyses; and review papers. One participant noted several ways components could be used: “In research (meta-analyses, review of relevant data on specific topics, and guidance in table/figure design).” (Post Doc, Ecology) This type of use conforms to conventional notions of how scientists incorporate previously published work into their own new work. What is different here is that a journal article component retrieval system provides a new method of accessing the journal
literature: more detailed, in-depth indexing is applied to individual tables and figures, which allows searchers to locate information of interest even if the entire article is not on that topic. A post doc in ecology noted that finding a figure using the prototype piqued his interest in an article he was sure he had obtained earlier: “Hmm, this is actually an issue I read before. Somehow I managed to miss the paper that I’m now interested in. I actually have this article, and yet I don’t remember this figure.”

In another common practice, researchers incorporate components into text or multimedia documents used to support performative activities, such as presentations in the classroom, at conferences, or at job interviews. Researchers want to illustrate concepts, organisms, and methods with graphics from the current research literature by extracting a high quality image from a published paper and placing it into standard presentation software. A journal article component retrieval system lets the researcher find a relevant object directly, simplifying the current practice of searching for abstracts, reading abstracts, selecting and obtaining the most promising articles, and only then assessing the relevance of specific tables and figures. To “…be able to drag it [table or figure] onto PowerPoint and make … class notes or something like that. That would be incredibly convenient.” (Post doc, Geology) In some cases, particularly for teaching, scientists are looking for particular figures and tables they have seen before: they often recall particular objects or their characteristics, but not the title, author or source of the corresponding article. A system supporting direct searching for figures and tables promises to make the process of locating them much more efficient. A professor of biochemistry uses extracted journal components to “incorporate current publications into classroom lecture.” One motivation for incorporating components from articles into presentations is that “in teaching, visual representation is very important.” (Post Doc, Biology)

Participants also identified ways journal article components can be used for making a variety of comparisons, a use not before noted by other researchers and thus one of the most interesting findings from this study. Researchers compare their own work, results, methods, or instruments to the work, results, methods, or instruments of other researchers, either to see if their own research is credible and on-track or to see if “someone has generated data that might be similar or complementary to my own data.” (Professor, Biology) A post doc in ecology explained how he used the prototype to look for and find photographs of microcosms because he was interested in building some to support his own work with fungi. In other cases, researchers were interested in finding published data for comparison: “Finding data (tabular or graphical) that met some criteria in order to compare with models.” (Post Doc, Biology) “… to build context & make valuable data comparisons [between] your work and other scientists’ results.” (Post Doc, Biology)

Finally, participants sometimes mentioned how journal article components could support the development of new intermediate objects that might later

be reaggregated into or support the creation of other fixed text documents, such as research proposals, journal or conference papers, software, etc. Examples cited by the participants included developing new formulae, models, simulations, data compilations and hypotheses. “Having explicit presentations of data, such as in table form, allows me to quickly use that paper for downstream applications such as formula generation or system modeling.” (Student, Biology) Another participant noted that “It would be very useful when trying to make compilations of existing data (especially tables).” (Post Doc, Geology) One professor of geology mentioned using “figures for modification and addition in research.” A post-doc in biology said “Generating new data is expensive & time consuming. Finding existing relevant data is therefore valuable.”

Implications

These findings hold implications for practitioners as well as LIS and communications researchers, both in terms of opportunities for service delivery and opportunities for research into the relationships between documents and scientific work at new levels of granularity. For service providers – both librarians and content providers – there are opportunities to create and market services that bridge the gap between researcher practices related to journal article components and the ability of current systems to meet those needs. One gap is the mismatch between researcher needs and the services provided by search engine providers, such as Google Images, that fail to discriminate between images related to scholarly work and popular images. Improved, higher-quality indexing of journal article components, along with improved search and retrieval systems, can help scientists efficiently incorporate components into performative assemblages and data compilations, and support the many kinds of comparisons they make between their own work and the work of others. While this prototype exposed figures and tables to the end user, it is conceivable that providing direct access to additional journal article components, including section headings, table text, and the methods, introduction, and conclusion sections, would also be valuable to scientists (see Bishop, 1999; Stelmaszewska & Blandford, 2004).

These findings also suggest emerging opportunities to conduct research into scholarly communications focused on artifacts at finer levels of granularity. Now that systems supporting retrieval of fine-grained information objects are available, LIS research should re-engage with scholarly communities to understand how scientists value, search for, and use these newly exposed objects. Holistic, situated studies of component use can also help reveal the mutual construction of relationships between the capabilities of information systems and other dimensions of scientific work, such as community values and reward systems, and evolving work practices. The importance of easy and direct incorporation of figures and tables into presentation software for use in the classroom, presentations of research, and job interviews was revealed by the current study and reported in this paper. It may be that the

The centrality of this practice has not been previously identified because of the concurrent emergence of digital library systems that provide end-user access to high quality digital images and the widespread use of presentation software packages over the past decade. It is reasonable to expect continued co-evolution of information practices and information systems will provide additional opportunities for scholarly communications researchers. For example, will scientists develop new citation practices that link to discrete journal article components instead of their embedding journal articles (Lindquist, 1999)? Or will scientists begin to create and publish figures and tables outside of the context of traditional journal articles? Finally, how will the design and implementation of metadata systems take the disaggregation of journal articles into account so that existing and emerging standards fit and support evolving scientific practice?

Acknowledgements

This work was funded in part by ProQuest CSA, developer of the prototype index used by the scientists who participated in the evaluation reported here. Participants in the project included Donald W. King; Bobbie Suttles, Center for Information Studies; and Alison Connor and Kelli Williams, Graduate Assistants.

References


---

1 Individual, session-level transaction log data was not available from the prototype during this project.
2 The involvement of ProQuest CSA in the project was limited to identifying institutions likely to participate, putting the researchers in contact with a librarian at each institution, and commenting on first drafts of the data collection instruments. The librarians at each institution recruited the participants. The researchers were solely responsible for all data collection and analysis. ProQuest CSA agreed at the inception of the project to allow the researchers to disseminate results without any restrictions.