HOW DO THE RESEARCHERS UTILIZE THE ELECTRONIC LIBRARY?

Helle Lauridsen

Proquest CSA

UK

hlauridsen@csa.com

Abstract

A scientific paper is a written report describing original research results whose format has been defined by centuries of developing tradition, editorial practice, scientific ethics and the interplay with printing and publishing services. The result of this process is that virtually every scientific paper has a title, abstract, introduction, materials and methods, results and discussion.

It is in this WRITTEN report the abstracting and indexing for A&I databases is normally done, not in the primary research results as represented by the tables and figures in the article.

This method has worked well for over a century as scientific information has been fairly limited, but with the information explosion[1] of the past 20 years information retrieval by traditional indexing was only temporarily saved by the speed and ease of electronic searching. At present we have so much information at our fingertips that the question is not to find it, but to limit it to the most relevant material.

Figures and tables represent the distilled essence of research communicated in academic articles.
Although the analysis contained in the surrounding text is important, it is clear that researchers are eager to view the actual data collected, observed, or modelled to determine the article's relevance to their own work. The summary of data displayed in figures and tables is a highly valuable surrogate for the typically unavailable raw data sets..

The primary objective of a literature search is to locate information relevant to researchers' interests.

Neither traditional article-level indexing nor full text-level indexing where all text within a document is searchable can locate those publications which contain specific data of interest. By indexing the variables defined in tables and figures, researchers can find data with pinpoint accuracy.
For years this has been too difficult to do as many figures only appear as .jpg images and thus are closed for machine indexing. But two years ago the idea was revived by an innovative persons within the company CSA and now the technology had been developed to be able to do the actual indexing of the many figures and graphs in the articles.

The concept is that all tables and figures contained within an article are indexed. The number of records in a Tables & Figures Index is an order of magnitude greater than those contained in a typical abstracts database. In the database each record is being assigned one or more general categories reflecting the 'type' of data display (e.g. Photomicrograph, Histogram, Line Graph, Map of Study Site, and so on).
The figures are being Indexed – The primary terms enabling accurate searching:
a. Subject Indexing – Key variables presented in the figure or table
b. Geographic Indexing –A applicable geographic terms
c. Taxonomic Indexing – The Latin names of organisms will be included when appropriate, most
often consisting of the genus and species names, but will include broader categories when available (e.g. family, class, etc.).
d. Statistical Indexing – Any standard statistical term relevant to a particular data display (e.g.
Analysis of Covariance, ANCOVA, Simple Linear Regression, etc.)
e. Other Relevant Data – an indication of whether the table or figure contains either an empirical or theoretical predictive model

[1] In its first year of publication (1907) Chemical Abstracts contained a total of less than 12,000 abstracts. By contrast, Chemical Abstracts published a million abstracts just in 2006

And finally each record is linked back to the source journal article

The perceived Benefits of Searching Tables &Figures was that:

-Targeted searches could be constructed by employing figure-oriented searches allowing the researcher to save time and match retrieval to specific data contained in the article.
-Researchers could ensure that the study actually focused on a specific variable, rather than simply referring to it indirectly (i.e. from another publication).
-Categories of objects could easily be browsed allowing easy creation of visuals for conference presentations, teaching or seminars.

In depth Market Research by Carol Tenopir

But all this was theory and before launching this very large expensive project for real, CSA wanted to make sure that the idea was viable and in the spring of 2006 an in depth market research was initiated: to make it non-company related CSA asked a research team led by Professor Carol Tenopir, with Donald W. King, Dr.Robert Sandusky 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 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 quantitive and qualitive data. They allowed the team to study both predictive questions, such as how indexing of tables and figures might be used by scientists, and functional questions such as what type 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.

The Findings [Tenopir, et al. 2006]
The Research team wanted to understand the baseline experiences, skills, practices, and levels of use of digital / electronic journal articles and digital objects as well as the tools that currently exist to help researchers locate these articles and objects. They asked questions designed to (1) elicit information from all participants about current use of journal articles and electronic systems and (2) understand the experiences of the subset of participants who already had experience using various existing systems to search specifically for digital objects, like photographs, maps, figures, etc.
Part of the questionnaire given to all participants at the on-site briefing meeting included eight questions that sought to elicit information about researchers' current practices with regard to searching for and using journal articles. Questions focused on number of articles read; percentage of articles read that were obtained or viewed electronically; systems used for searching during the past 12 months; who performed the article searches; the participants' ratings on the importance of and their satisfaction with five general attributes

of search systems; and comments on favourable or unfavourable search experiences during the past year.
This line of questions helped situate the participants in terms of their general usage of digital abstracting & indexing systems and similar technologies.
A pair of questions asked the participants to reflect upon how many times they read scholarly journal articles in the past 30 days and what percentage of the articles they read were obtained / viewed in an electronic format. In the context of the first question, the term *reading* is defined as "going beyond the table of contents, title, and abstract to the body of the article." The most frequent response was 10-19 articles, reported by 22 participants (37%); another 37 participants (62%) reported reading between 10 and 30 articles in the past month.

The companion question asked the participants to indicate the percentage of the articles they had read in the past month that were obtained or viewed in an electronic format.

An overwhelming majority (39, or 65%) of the participants indicated 90% or more of those articles were viewed or obtained electronically, with 17 of the participants (28%) indicating that 100% of these articles were obtained electronically. An overwhelming majority of respondents (45, or 75%) report obtaining the vast majority of their articles electronically, when all participants reporting obtaining 70% or more of their articles online are combined. Another 4 participants (7%) reported obtaining or viewing from 20% to 69% of their articles electronically. A single participant reported obtaining no articles electronically. These particular responses demonstrate the near-universal dependence upon networked systems for access to scholarly articles in these disciplines.
We asked participants to tell us the approximate proportions of their online searches conducted during the past 12 months using specific kinds of resources. Resource types listed were *Web search engine* (Google, Google Scholar, Yahoo!, Alta Vista, etc.), *electronic indexing / abstracting service* (CSA, ISI Web of Science, BIOSIS, CAS, Academic Search Premier, PsychINFO, PubMed, etc.), *online journal collections* (Highwire, JSTOR, MUSE, etc.), *current awareness or alert service* (Current Contents, listservs/mailing lists, RSS feeds), and *other*. No participants reported using any one kind of resource exclusively.
Electronic indexing / abstracting services were the most frequently used kind of resource: 35 participants (58%) indicated that 60% or more of their searches, and 49 (82%) reported that 40% or more of their searches were performed using electronic indexing / abstracting services. One participant, however, reported never using this kind of resource (Figure 9). Web search engines were the next most cited kind of resource (Figure 10), but frequency of use was much lower than electronic indexing / abstracting services: 42 participants (70%) reported that 39% or fewer of their searches were conducted using Web search engines.

This wide use of the library's resources by researchers is the opposite of the picture given by speakers (Stephen Abrams, Sirsi Dynix, Tom Davy Thomson learning, to name a few) at recent library conferences where the Google/Google Scholar is shown by interviews and short video clips to have completely taken over from the library as the place to search information.

The information in the survey by Carol Tenopir is backed up by a couple of Danish Usability reports, in "The hybrid library: from the users' perspective"[Lomholt Akselbo, et al. 2006] it says: "Many of those interviewed – students as well as researchers – express a great deal of skepticism about Google, because the search engine simply cannot make nearly so much quality material available as the libraries can offer. In other words, the libraries have a very strong standing in relation to Google when the users in the study reach the stage of retrieving full-text material – printed as well as electronic."

However in another usability test "Det brugervenlige digitale forskningsbibliotek"[Sandberg Madsen, et al. 2004] the test shows that users have problems

using the digital library: to decide which databases to use, to find the full text articles and to quickly and efficiently evaluate the information in the bibliographic databases.

This problem is increased by the mere nature of electronic information search: users rarely go to the physical library but most often conduct their searches from their offices or their homes, thus making personal, professional help very difficult[Sandberg Madsen, et al. 2005] .

Many libraries have tried to work with this problem but the innovations from database provider side have been few and far between: The launch of the Citation Indexes by Eugene Garfield by adding cited references to traditional bibliographic indexing records about scholarly articles [Jacso. 2007]helped pointing at the most cited articles within a given subject and it has been adapted by many major databases since. But not much has surfaced since.

Conclusion from initial surveys.
In other words the students and researchers are happy to use their libraries' electronic resources but it can sometimes be very difficult to find relevant, precise information in a timely and accurate fashion.

So there was a need for innovative ways of drilling into the core information – but would CSA's perceived idea about a tables and figures index meet the researchers' needs? The second part of the Carol Tenopir White Paper[Tenopir, et al. 2006] looked into this problem, and the findings were pretty conclusive:

One of the questions asked, was designed to elicit information about researcher expectations of a system providing high-quality search for objects drawn from scientific articles prior to their exposure to the Tables and Figures index prototype. The participants were asked to describe situations in which the ability to search for and retrieve objects such as tables, figures, graphs, maps, and photographs would be valuable to them. A wide range of responses was received from all 60 of the participants and their responses can be grouped into the following categories:

- Teaching, lectures, talks, presentations including incorporating tables and figures found directly into presentation software, such as PowerPoint
- Locating and retrieving data of particular types, such as tables, graphs, figures, maps and photographs
- Making comparisons between one's own work and the work of others as well as comparing the work of multiple other researchers for a variety of purposes; putting one's work into the context of research in the discipline
- Gaining faster and more precise understanding of the work reported in other papers by direct examination of the objects embedded in other articles
- In support of writing and other forms of scholarly production including conducting meta-analyses and writing review papers, writing journal articles, writing research proposals, developing formulae and models, and generating hypotheses
- Faster and more efficient searching, with smaller, more precise results sets
- Supporting the transformation of practice and supporting learning, by researchers, of new skills and methods, including how to effectively present scientific results in tables, figures, and graphs
- Librarians noted the utility of locating objects directly to answer reference questions
- Concerns expressing doubt or concern regarding the capability to search for objects drawn from scientific articles
- Ideas for useful capabilities that are beyond the scope of the current Tables and Figures index prototype

The most frequently identified situations where searching on tables, figures, graphs, etc.would be valuable had to do with preparing for lectures, talks, and presentations – both in the classroom and to other audiences, such as at conferences or job interviews. When

researchers have a specific point to illustrate in a lecture or presentation, they want to be able to find a relevant object without reading abstracts, then obtaining the most promising articles, and then examining and extracting the tables and figures. In some cases, particularly relating to teaching, they 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. Direct searching for figures and tables has promise to make that process more efficient.

A follow-up open-ended questionnaire prompted the respondents to elaborate on how this capability made a difference in their search and discovery process. Overwhelmingly, participants alluded to the fact that this capability saved time and provided quicker access to information. “I can find the tables and figures that I need quickly, [and] it can save me a lot of time. I can work more efficiently” (Post Doc, Biology). One participant mentioned the increased efficiency of the search process, stating “It makes the search much quicker when it is focused” (Post Doc, Biology), and another noted that “the tables and figures are really helpful for scanning large sets of data first” (Post Doc, Oceanography). Some participants specifically noted that this quicker access and search time was a convenient aid to presentation preparation: “[i]t takes less time to find the information I want and especially I would find this useful when making a presentation” (Student, Biology). Another wrote: “I could find relevant information more quickly and images that were useful for presentations and research” (Professor, Engineering).

The findings above are just a small part of the feedback we got from the librarians and the researchers of the study, some very positive: Participants had favorable comments regarding the stand-alone figures and thumbnails.
One enjoyed “being able to view the demographic information on the subjects in a study prior to or without getting the full article” (Librarian). Similarly, another felt “it makes it more interesting to look at the figure first; you are more inclined to read about it…than [from] looking at the abstract first” (Post Doc, Biology) One participant “enjoyed the ability to see the figures and tables within a search topic. I feel that the thumbnails are often sufficient, that once I see those I know whether or not the item is worth pursuing” (Post Doc, Geology).
But not all was favorable: Many participants commented on problems related to images and thumbnails. The prototype had some problems with the images not enlarging and some of the figures were too small and of too poor quality to be of any use. The quality of the captioning could also be improved and it was in general agreed that the entire caption was important and should be included.

During the fall of 2006 the entire prototype was taken to pieces: the images needed to become MUCH better, the captions clearer, response times to be kept at a minimum. Smaller thumbnails appeared in the search results

and easy mouseover on the figures in the actual record provides the required complete caption:

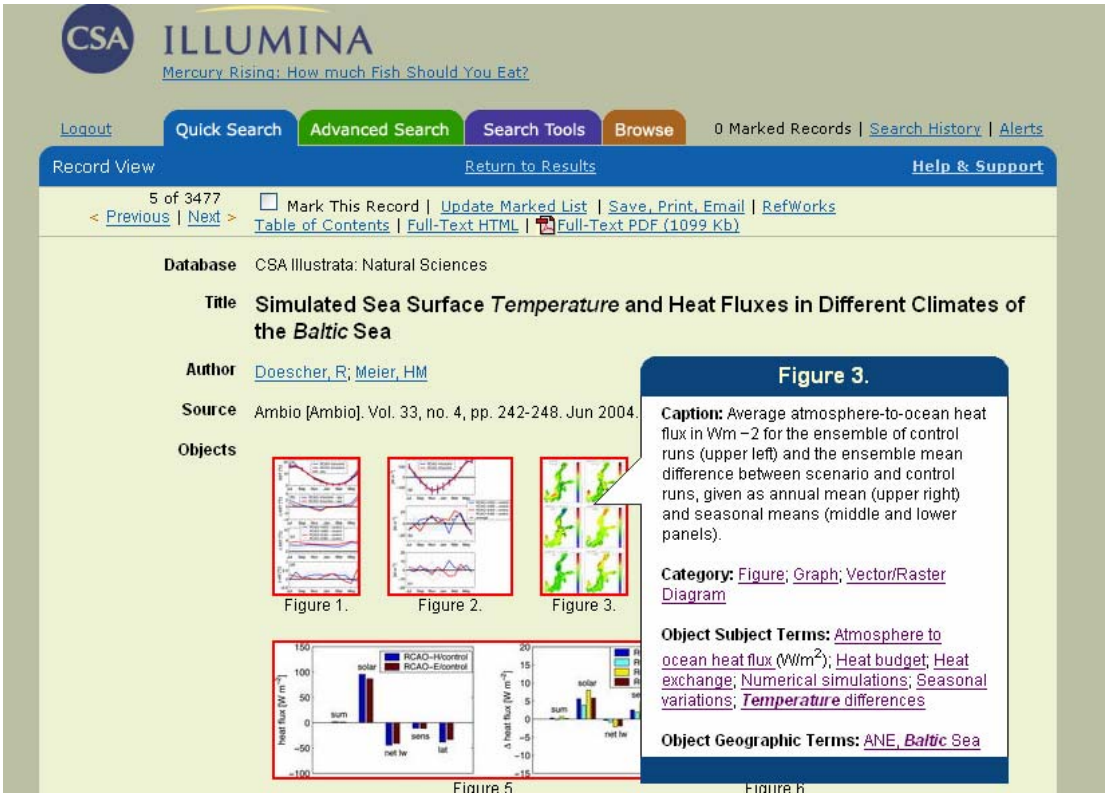
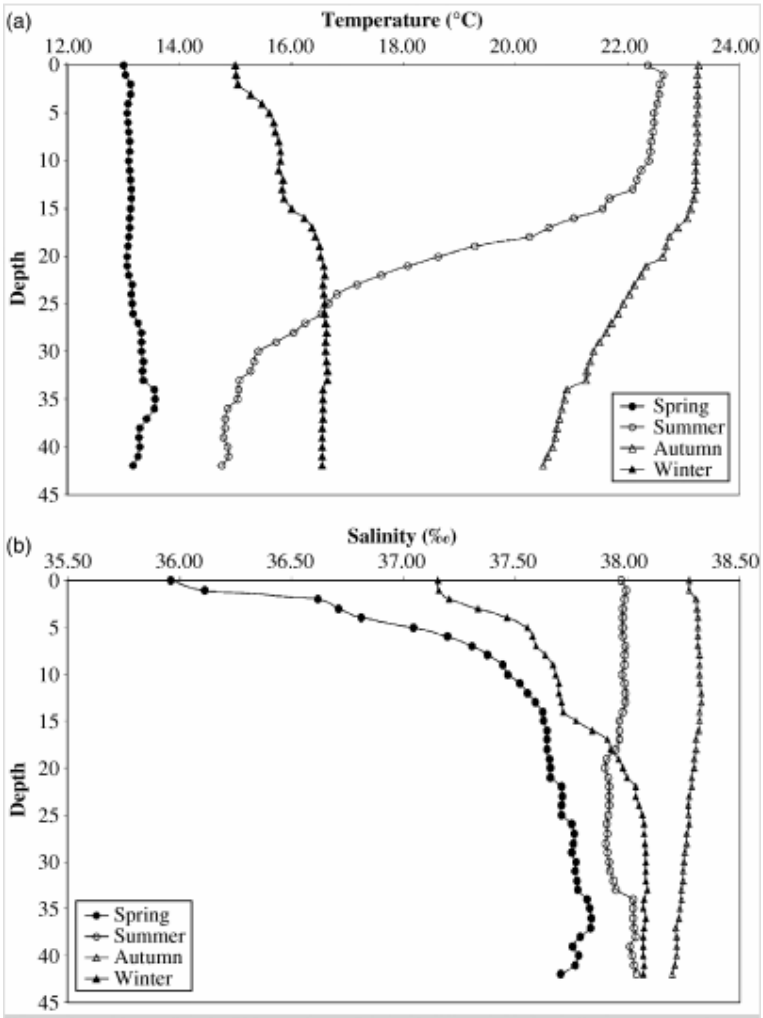


Figure 1: screenshot from CSA Illustrata showing mouse over caption

It was released under the name CSA Illustrata – Natural Sciences in January 2007
The publishers in the process of signing on for having their material indexed now cover almost all the large Academic publishers and the number of records in the database has now reached a critical mass of over 1.000.000. Beyond indexing the text of the captions of illustrations, all illustrations (objects) are enhanced and indexed by descriptors, subject terms, and digital object identifiers. Where applicable, taxonomic, geographic, and statistical terms are added. Each is searchable.

Since then many researchers and librarians have been able to test the new indexing method and the feedback has been extremely overwhelming.
Firstly the already perceived benefits of being able to search in data not hitherto indexed and thus retrieving hidden information, is holding true, at a demonstration a researcher commented that it was impossible to find any information about the temperature in the Ligurian sea – a search was launched and a few seconds later a table giving exactly this but published in a paper on feed pellets in Mediterranean water was found.



Vassallo, P., Doglioli, A. M., Rinaldi, F., & Beiso, I. (2006). Determination of physical behaviour of feed pellets in Mediterranean water [Figure 1]. Aquaculture Research, 37, 119-126.
Publisher: Blackwell Publishing Ltd.

Figure 2: indexing of information not formerly available

The rapidity with which you can find relevant illustrative information impresses most people seeing the database for the first time and the human brain and it's way of processing information should not be discounted in this matter: Only about 20% of the worlds population learns best from text based information2, the rest from one of the other 4 learning styles – one of the most common of these being the image based.
Many people find it much easier to shift through search results including images and locate the relevant article fast and efficient

[2] Stephen Abrams, UKSG conference 2007

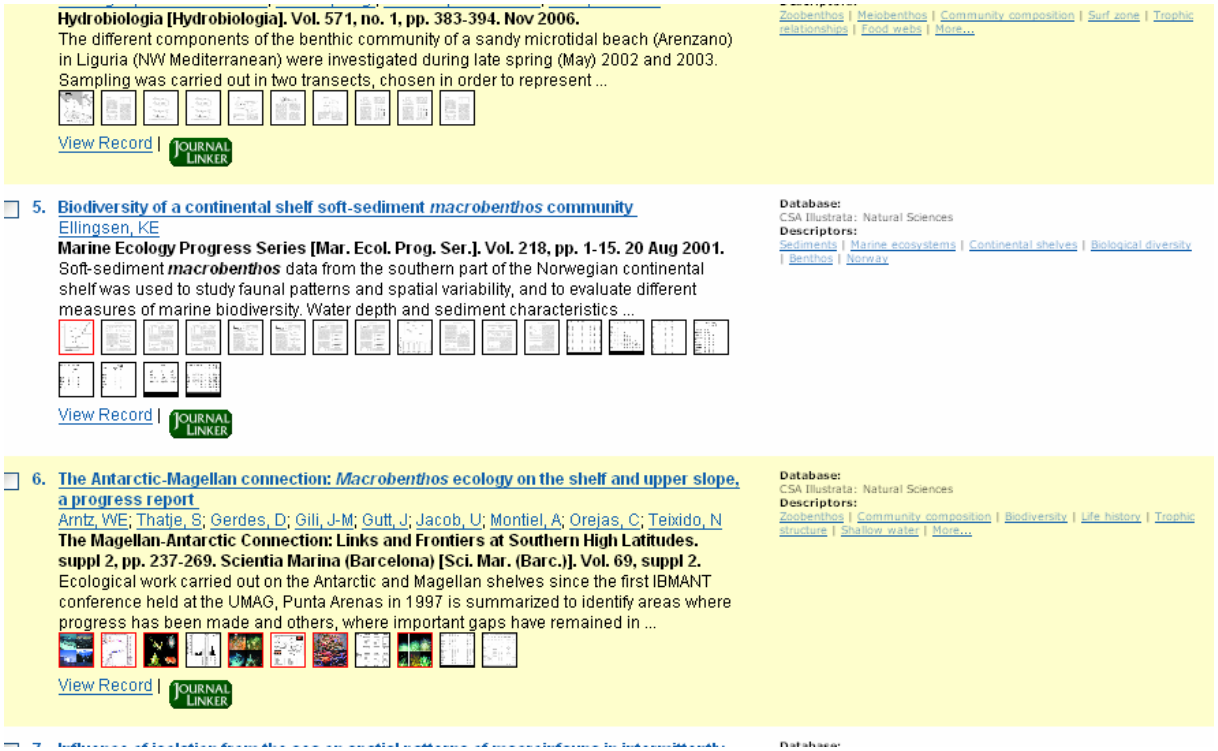


Figure 3: screen shot from CSA Illustrata showing small marked images, easy to navigate by the human eye

Very often a database search is for an extension of already present knowledge and a quick view of the article images will easily indicate if the article has any interest.

This method giving a quick visual overview of the article content is also used by the researchers:

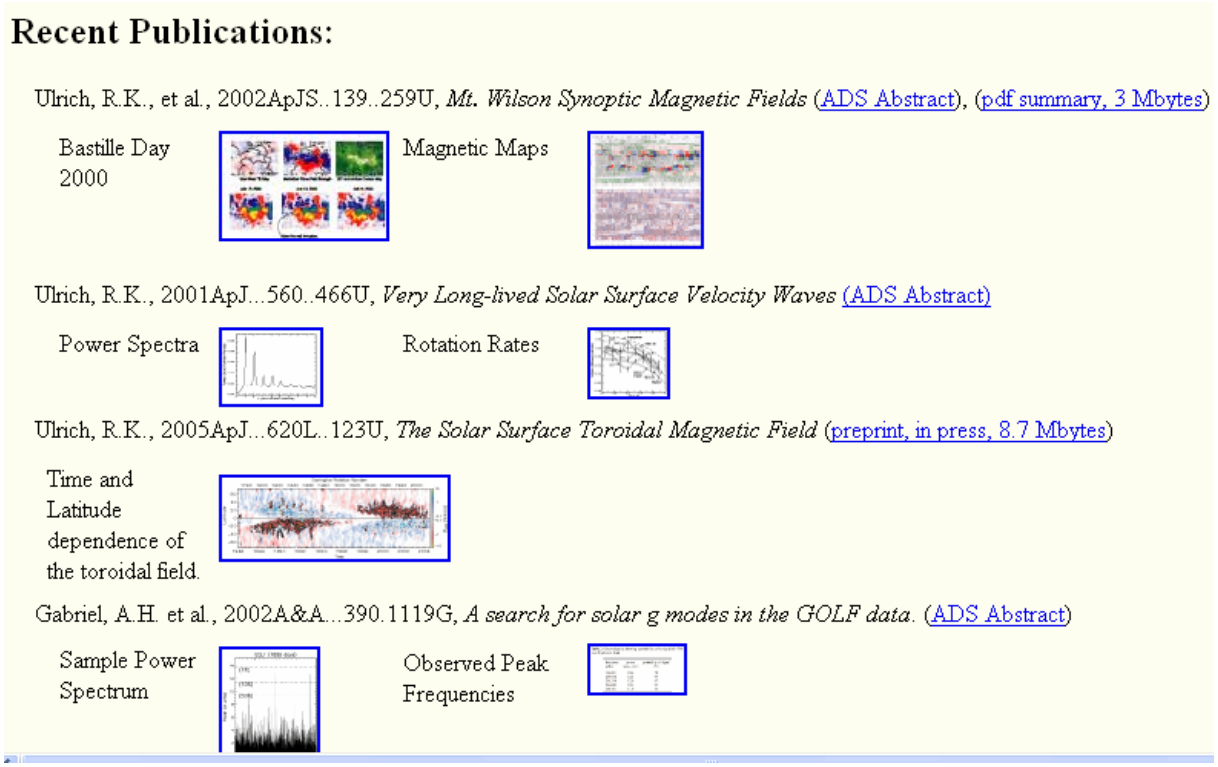


Figure 4: Screen print of researcher's webpage

Conclusion:

The idea of indexing article illustrations can be difficult to grasp and takes some time getting used to. But the need for more precise searches, for more exact ways of utilizing the masses of Electronic material most modern Academic and Research libraries invest in, for helping the users find material in other ways than text based is apparent and well known by the library community. Hopefully this new way of indexing will encourage both more precise searches as well as the serendipity that disappeared along with the browsing of new journals in the soft chairs of the print journal reading room .

References:

TENOPIR, C., SANDUSKY, R.J. AND CASADO, M.M. 2006. The Value of CSA deep indexing for Researchers. *White Paper* .

LOMHOLT AKSELBO, J., ARNFRED, L., BARFORT, S., BAY, G., BAGGER CHRISTIANSEN, T., HOFMAN HANSEN, J., TVERMOES JENSEN, H., BACH MARKUSSEN, G., MORTHORST, A.M. AND POLTORAK NIELSEN, M. 2006. The hybrid library: from the users' perspective. A report for the DEFF project "The loaners' expectations and demands for the hybrid library". .

SANDBERG MADSEN, J. AND GARDNER, J. 2004. Det brugervenlige digitale forskningsbibliotek, best practice rapport baseret på usability test af 11 store forskningsbibliotekers websteder.

SANDBERG MADSEN, J. AND SØNDERGÅRD, P. 2005. **Det brugervenlige digitale forskningsbibliotek**. *DF Revy* 28, 12.

JACSO, P. 2007. CSA Illustrata by Jacso. *Online* 32,