The Electronic Resources Project of the Faculty of Information Studies

Marte Misiek and Gerry Oxford
Faculty of Information Studies
University of Toronto
misiek@fis.utoronto.ca
oxford@fis.utoronto.ca

The main goal of the Electronic Resources Project is to improve access to our electronic journal collection. From our Web-based search form, users can do Boolean searches of full text, title, author, and other categories on a per-article basis, and then select and retrieve articles from a list of hits. Search terms are highlighted within the text. We achieve this result by marking up journal issues using SGML, loading the issues into an OpenText database, and providing for searching through the Web form. Our electronic journal collection of about thirty titles is also catalogued and searchable through the Web interface to UTCat, the University of Toronto union catalogue, and accessible through hot links in the 856 fields of their bibliographic records.

1. Introduction

Access to scholarly literature within libraries exists on three different levels. At the most basic level there is the item itself. Large collections provide scholars access at this most basic level. The middle level is access to the contents of the item, provided by book indexes or abstracting and indexing tools for journals. Most often libraries utilize work done by others at this level. The top level, access to an item within the context of the body of scholarly literature, has traditionally been handled by cataloguing the item so that it can be located in relation to the collection as a whole. Electronic journals are a recent but rapidly growing form of publication in virtually every academic discipline. As of May 1995, there were nearly 700 electronic serials, a sixfold increase over the number listed in 1991 (Okerson 1995, i). If these journals continue to develop as a viable means of reporting scholarly research, it will be necessary for future electronic libraries to incorporate them into their collections at all three levels.

2. Nature of electronic journals

Electronic journals are similar to paper journals in many ways, including their publication types (newsletters, journals, etc.), volume and issue numbering and dating, and publication frequency and regularity. A growing number are peer-

reviewed. Some electronic journals are issued only in electronic format; others began on paper format and are now issued in both. Electronic formats vary greatly. Initially many were issued as plain ASCII text; now many are also marked up, or only marked up, in HTML or in other ways for better electronic display. Physically, some consist of one long file; others come as one file per section or article with a separate file for a table of contents and/or index. Of course, the integrity of issues is not assured now that many journals are available as hypertext documents with links to other documents or sites.

Subscriptions to these serials are obtained electronically via e-mail over the Internet. Many are free; however, there is a small but growing number that are priced. The method of delivery also varies. Many come automatically to a subscriber's e-mail address; however, for some, a notice is sent to subscribers when the next issue is available and then, if it is wanted, the subscriber must request it by e-mail or retrieve it by FTP.

3. Traditional treatment of electronic journals in libraries

At the bottom level of the collection, several university libraries have been collecting electronic serials for several years. North Carolina State University (NCSU) and CICNet in the U.S. midwest are two of the more important sites that have large electronic journal collections. Some collections are "real" in the sense that the journal holdings are physically present at these libraries; others are "virtual" since they point to journals held elsewhere. A review of holdings in our subject area reveals several problems with relying on the holdings of other collections. The location, or URL, of the collection may not be stable or maintained; many libraries continued to point to the NCSU gopher for holdings after holdings were being updated only in their new Web location. Some still point to an old CICNet URL even though it changed over a year ago. Any particular title is often not entered consistently or uniquely at a site. A title may appear in two to four different locations at a single site. Each journal entry is often not standardized, and its holdings may vary from one entry to another. Another problem is that the holdings are not always being kept up to date; new issues seem to be added to the collections in batch mode, so the holdings may be updated only every three to six months.

At the middle level of content, access is currently very basic. Most of the electronic journals are not yet indexed in the customary indexing and abstracting tools. Some electronic journal collections provide a WAIS-type search capability for access to the contents themselves. WAIS allows a fairly primitive full-text search by keyword or phrase of an indexed database. An example of the results of WAIS search on the phrase "SGML" in the electronic journal *The*

Public-Access Computer Systems Review (*The PACS Review*) at NCSU is included in Figure 1.

Retrieved hits are presented as a list of the file names that match the search. These files could be complete issues, articles or even a table of contents or index, depending on the format of the journal searched. The file names are usually not indicative of subject content and are arranged in no apparent order. There may be duplication of hits if the contents and article are in separate files. There is no indication of the number of items in the hit list. A hit may not be relevant, since the search terms may be in different articles in an issue. Search terms are not located within the text; it may be necessary to browse through many pages of text to determine if the retrieved items are of use.
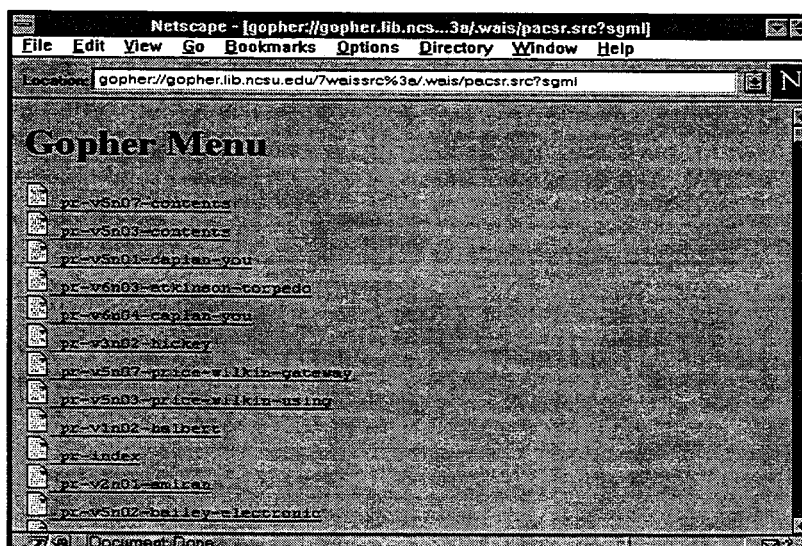

Figure 1. Results of WAIS search, NCSU

Another, less useful, access to contents is provided by a Web search, such as the kind offered by Yahoo or OpenText, by subject keywords or phrases or authors' names. The results of these searches are even more problematic in terms of likely relevance and ease of determining usefulness.

At the top level of knowledge context, access is at most basic, if not non-existent. Currently most electronic journal collections are maintained separately on gopher or Web servers. Often you must know of the existence of

a collection or stumble upon it through a search of the Internet. The cataloguing of these journals to integrate them with the body of paper-based scholarly literature is still rarely done.

4. Electronic Resources Project

4.1. The proposal
In February 1994, we determined that no work was being done in building collections of electronic journals in our subject area or in providing improved access to them along the lines of CD-ROM/online-type capabilities. In May 1994, the Electronic Resources Project received funding for a period of one year from the Academic Priorities Fund of the University of Toronto. A major phase of the project involved developing a collection of electronic journals in the area of library, information, and archival sciences, working out procedures for incorporating the continuing maintenance of the collection into the normal functioning of the Faculty of Information Studies (FIS) Library with as little human intervention as possible, and improving access to the collection.

At the beginning of the project, we concentrated on providing access to electronic journals at the basic level and at the middle level of the contents of the collection. As the project progressed, it became apparent that we needed access at the top level of the catalogue also. Throughout we have been working toward providing seamless access to the three levels of scholarly literature in electronic format from the World Wide Web.

4.2. The bottom level—the collection
Our preliminary research discovered no specialized collections in our subject area. A list of prospective titles was developed from collections of relevant titles at NCSU and CICNet, the *Directory of Electronic Journals, Newsletters and Academic Discussion Lists* (Okerson 1994), and the Internet list NewJour. Selection of the titles was made according to the Library's *Collection Development Policy* (University of Toronto 1994) in the same way selection is done for our paper journal collection. The resulting list included about thirty to forty relevant titles.

In May 1994 when we began our project, most electronic journal holdings were kept on gopher servers; none of our publishers maintained access to their own journal's holdings by anything other than FTP. Relevant journals held as part of existing electronic journal collections often had the previously discussed problem of not being current. We decided to amass a physical collection of our selected titles at FIS, especially since our plan for improved access would probably require in-house access to the physical issues. This collection is

available on both the FIS gopher and Web. Our work in the area of processing incoming journals was based on work done by Eric Lease Morgan at NCSU (Morgan 1995).

4.3. The middle level—the Web-based search form

4.3.1. Early work

Improvements in searchability along the lines of CD-ROM and online-type searching without major expenditures of staff time were our biggest challenge. Among promising early leads, we hoped that a WAIS-type search that allowed searching by fields would be available. The extension of the Z39.50 standard into the area of non-bibliographic data also seemed promising. We were unable to develop anything from these leads. Finally, we settled on our current WWW/SGML/OpenText solution to achieve improved searching capabilities for middle level access.

4.3.2. Hardware/software environment

The database engine we use for searching is OpenText 5, from the OpenText Corporation of Waterloo, Ont. The software runs on a Unix SunSparcstation and supports full-featured search and retrieval on databases marked up using SGML-compliant markup languages. It uses the PAT technology (OpenText 1994) to create an index based on suffix arrays, which allows for the searching of very large text databases with consistently fast response times. It also supports Boolean and proximity searching capability. The software is used as the search engine for the OED and also as a WWW-search tool.

4.3.3. SGML/DTD

The journal issues themselves have been marked up in an SGML-compliant language. SGML is an ISO standard (ISO 8879:1986) that is a grammar or set of rules for defining markup languages. HTML is an example of an SGML-compliant language. SGML compliancy is useful because it is becoming a common electronic publishing tool. Also, text marked up in an SGML-compliant language can be easily and automatically translated to other SGML-compliant languages (for example to HTML for display on the WWW).

Many suggest that it is not worth doing more than HTML markup of electronic journals. It is true that HTML is an SGML-compliant language, but its tags indicate only how the document is to be presented for display; they do not indicate document structure. Using a richer, structure-oriented tagging system for content would allow us to achieve some of our goals with respect to field and article searchability and also retrievability of individual articles

instead of complete files. Markup of the fields needed for citation-type entries would allow us to improve the presentation of entries for our hit list with respect to the content of the hit list entries and also to utilize HTML for Web display by translating to HTML. An added benefit from the use of a standard is that the marked-up texts will then be portable and reusable as other database search engines and display conventions become available. This parallels the use of the MARC standard for the electronic storage of bibliographic information.

A Document Type Definition, or DTD, is a markup language or set of rules that defines the structure of a particular class of documents and adheres to SGML standards. Two were of particular interest. ISO 12083 (International Organization for Standardization 1994) has a section dealing with journals. We also examined the Text Encoding Initiative's work (Sperberg-McQueen and Burnaud 1994), which is a set of guidelines for the application of SGML to literary texts in the humanities. Both standards were too complex in terms of both the number of tags defined and the complicated embedded structure of the tags. A complex structure may be useful at the document creation stage; however, our texts are pre-existing and in a wide variety of formats. We have no control over their structures and are required by participating publishers not to tamper with them. Also, at least one critical element, special issue editor, was not included in either standard.

It became obvious that we needed to design our own DTD to meet a number of requirements. Specific elements such as author and title were needed to enhance searchability. The number of actual tags should be kept to the minimum needed to enhance either searchability or presentation of the texts in order to reduce markup time. Our goal was to automate as much of the markup as possible and to minimize staff editing time. As we indicated earlier, our electronic journals followed no standardized format. Our DTD needed to be flexible and simple enough to accommodate this lack of standardization without changes to the text. Discussions with some of our publishers indicated that many required that the displayed text be identical to the published ASCII version as a precondition to participation of their journals in our project.

As we gained experience with the multiple journal formats, it became apparent that an automated markup would be time-consuming. We settled on using a template for what we call the front matter (FISFRONT), which is appended to the beginning of each issue. This section, as shown in our DTD in Figure 2, contains information that either rarely changes (such as the journal title and editor, ISSN, copyright information, and FIS markup note) or changes only minimally from one issue to the next (such as volume and issue numbers, issue dates, and issue URL).

The number of tags added to each article is minimal: basically author, title,

```
FISJRNL             electronic journal
   FISFRONT         front matter - automatically generated
      FISSTITL      serial title (as taken from cataloguing record)
      FISATITL      alternative title (as taken from cataloguing record)
      FISEDITR      editor of journal
      FISCPYRT      copyright information
      FISSPED       special editor
      FISCDATE      issue date (as taken from issue itself)
      FISSDATE      sort date (created in the form year mo day, 19950130)
      FISVOL        volume number
      FISISSUE      issue number
      FISISSN       issn
      FISREPR       reprint information
      FISNOTE       FIS mark-up note
      FISNAME       standardized journal name (user name for UNIX)
      FISRECD       date received
      FISEJURL      URL of full journal issue at FIS; must be changed for each issue
   /FISFRONT
   FISRTEXT            retrieved text (complete article)
      FISARTCL        searchable part of article
         FISAUTHR     authors (in the form lastname, firstname, lastname, firstname)
         FISTITLE     article title (omit leading article and quotation marks)
         FISKEYWD     keyword (not currently used)
         FISABS       abstract (if at beginning of article)
      /FISARTCL
   /FISRTEXT

or

   FISRTEXT
      FISREV          article which is a review
         FISAUTHR     author of review
         FISTITLE     title of review (may be the same as reviewed title; if it is add (review) after title)
         FISRAUTH     reviewed author
         FISRTITL     reviewed title
         FISKEYWD     keyword
         FISABS         abstract (if at beginning of article)
      /FISREV
   /FISRTEXT

/FISJRNL
```

Figure 2. FIS electronic journal (FISEJRNL) DTD

keywords, and abstract. These tags reflect the kinds of searches we want to be able to do. A slightly different set of tags is provided for review articles to allow the indication of the author and title of the item reviewed and to provide the capability of searching these articles as a separate collection.

Our DTD allows us to break each single file issue into its separate articles so that the search, retrieval, and display of matches is always done at the article level and never at the issue level. This approach differs from the WAIS search approach, which always searches and retrieves at the file level.

We also reconstruct each multiple file issue into a single file and omit the contents file from the text to be searched. In a WAIS-type search, searching the contents files as well may lead to duplicate hits since both the contents entry and the article may match the search request. We join the contents to the articles to reduce this kind of duplication but still provide access at the article level.

Some modifications were made for retrieval purposes. In order to make the author sort meaningful and the display more consistent, we re-enter the authors' names in the form *last name, first name*. A meaningful title sort is achieved by leaving out the initial *The*, *A*, or *An* of the marked-up title. To overcome the lack of standardization of issue dates and produce a meaningful date sort, we have two date tags—FISCDATE, for the date as taken from the issue itself, and FISSDATE, which is a standardized sort date in the form *year/month/day*. A standardized journal title tag, FISNAME, allows us to provide for searching a serial throughout title changes.

As we began running test searches, we discovered that some resulted in hits on every article in the database, even though the articles had nothing to do with the word being searched (the word *copyright* is an example). Some of these meaningless hits were the result of such information as a copyright statement, which is often carried at the end of every article. This realization led to the creation of two separate tags to distinguish between the text of the complete article to be displayed and the searchable part of the article.

4.3.4. The collection

The journals in our in-house collection are all free. When we began the project, we felt we were not violating any copyright restrictions by collecting them. However, once we decided to mark up the journals in SGML, it became apparent that we must contact the editors/publishers to explain our project and request permission to use their journals. We received permission from sixteen; one publisher refused; the rest did not respond. In the prototyping phase of the project, we felt we had enough high-quality respondents and did not follow up our initial request with further letters. A decision was made to start with the

journals having the most scholarly content. *The PACS Review* was our test journal. Four more journals have now been added to the collection.

4.3.5. Processing incoming journals

Initially, we hoped to be able to do much of the processing of incoming journal articles automatically, by having Perl scripts on our Unix server recognize the issues as they arrived via e-mail, parse them for their generic parts (headers, dates, volume and issue numbers, etc.), and insert the appropriate SGML. The script would then forward the partially marked journals on for manual correction as well as marking up of the actual content of the journal (articles, reviews, authors, etc.).

Writing scripts was time-consuming and required very specialized knowledge. When customizing a script for each subscribed journal, the programmer had to code flexibly enough to handle the minor variations in header format that tended to occur between issues, but not so flexibly that the processing produced garbage. Many journals, moreover, made substantive changes to their look fairly regularly, which then involved a programmer's rehashing the processing script yet again. A simpler solution has the markup person place a standard header file at the beginning of each new journal issue.

4.3.6. Tools and procedures for marking up and parsing

Rather than try to automate the markup, we concentrated on providing the markup person with better tools. We began by considering a number of software vendors' SGML-authoring tools, such as Author/Editor from SoftQuad, but were unable to find any that were suitable. In late 1994, authoring programs were written to help users compose new work with SGML, rather than convert a pre-existent text into SGML. Most programs lacked even the ability to import ASCII text. They all worked on the paradigm of "give us the DTD, we'll give you the structure to fill in," rather than helping the user add structure to text already imported.

Our needs, ultimately, were rather simple: the markup person would insert a standard header for the issue, modify the header, and then insert some additional tags into the body of the document. The list of tags we needed to support was not very long. We elected to write some macros using WordPerfect for Windows, and then create a special WordPerfect toolbar with those macros. Figure 3 contains a copy of our toolbar. Thus, for example, the markup person can highlight the title of the article within the body of the text and then click on the <FISTITLE> button in the toolbar, in order to get FISTITLE start and end tags placed around the text. This method has worked quite well and allows us to mark up an issue quickly.
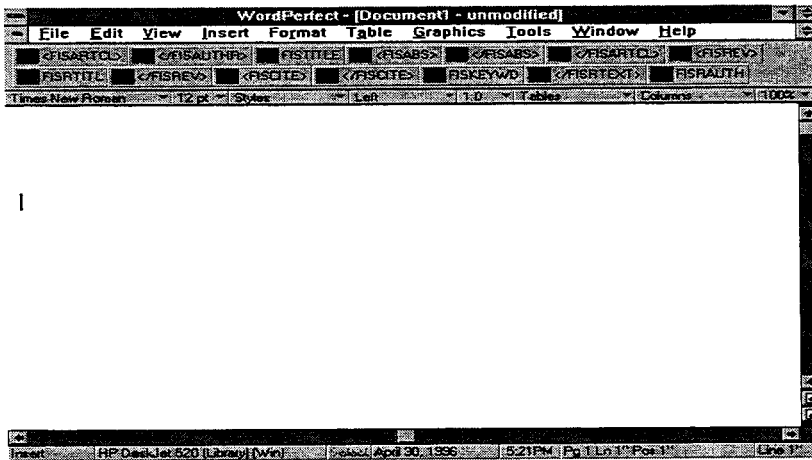
Figure 3. WordPerfect toolbar for markup

Of course, WordPerfect does not parse SGML documents to see if they conform to a DTD. We developed a procedure for parsing the documents and correcting problems with the document markup. When the journal issue has been marked up and successfully parsed, it is moved to the appropriate directory on the server, and the database of SGML documents is recreated.

4.3.7. Web-based search and retrieval access

Although an OpenText database can be queried by means of a specialized powerful client software called TextQuery (also from OpenText), we elected to use the Web as our distribution medium, both for its wide availabilty and the familiarity of the tools to our audience. Our markup scheme was not so complex that it couldn't be adequately addressed with the simple forms mechanisms available on Web browsers.

A Web-based search form, shown in Figure 4, allows access to the contents of our collection from the FIS Web pages and through the UTL Web OPAC interface. The search form allows for keyword or phrase searching in a number of fields including author, title, and full text. Boolean search operations and proximity searching are provided. The number of hits for a search is indicated, and it is possible to narrow the search further. The search and retrieval is performed at the article level to avoid duplication and improve appropriateness of hits. It is possible to sort the hit list in several ways including by date, journal, author, or title. An example of our improved hit list display for a search on the

Figure 4. FIS Web-based search screen

phrase "SGML" in *The PACS Review* is shown in Figure 5. It includes full citations for articles. Search terms are located and highlighted within the text to assist the user in determining relevancy. It is possible to search on a single journal or the entire collection of five titles. Some help buttons are included.

In order to search for articles in the collection, a user must connect to the URL for our search screen and enter the parameters of the search. Specially written programs on our UNIX server translate the search parameters into actual search commands for the database, execute the search, and then return the results of the search to the user's screen, translating from our native DTD to HTML on the fly. A similar process occurs when a user requests a particular article instead of just the results of a search.

The CGI script invoked by the form is written in Perl, and is fairly complex. Its job is twofold:
- Convert the field values from the filled-in form into a sequence of commands to the database engine that faithfully replicate the logic of the search as represented by the form;
- Convert the result set (returned by OpenText as an ASCII text file with SGML markups intact) into HTML-marked text to be returned to the
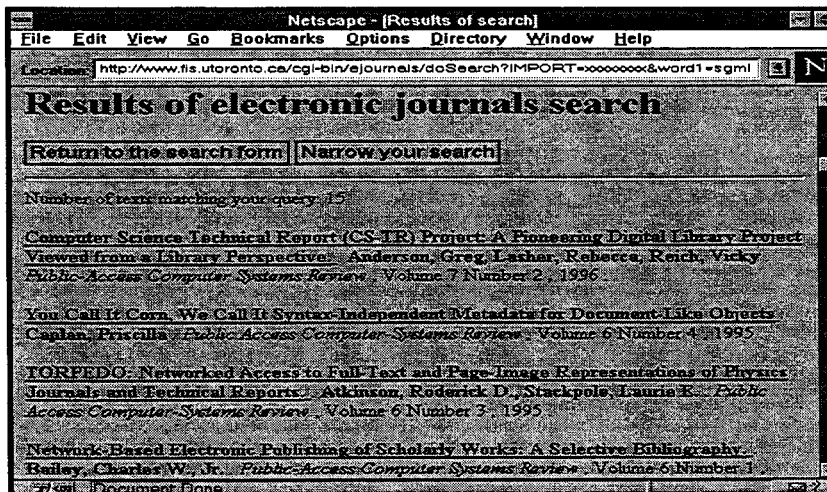
Figure 5. Hitlist display from FIS e-journal search

user's computer screen for displaying via the browser program. The script reformats the list of hits into a standardized bibliographic display, and inserts links around the titles of the articles that lead to the article itself.

A second Perl script is invoked when a user selects an article to view from the returned list of hits. This script executes another search of the database, using the link created by the first script.

Finally, a third script handles the problem of composing a new combined query if users should choose to narrow their search, an option available to them on the Results screen if their query resulted in ten or more hits.

In order to minimize the user's wait-time, the conversion to HTML is kept as simple as possible; where we could, we used one-to-one mappings between the tags of our own DTD and HTML. This approach placed some constraints on us in terms of the final look of the hit lists and articles; we didn't want to re-order elements of the texts, for example. Since we were already committed, however, to reproducing the text of articles exactly as they had been published, this was not really a new burden. We were able to add some basic enhancements to the article without substantively changing its look or increasing the processing time: useful header and trailer bibliographic information; a link to the full issue from which the article had been drawn; and a series of internal links within

the retrieved document to all occurrences of a found search term, to help the user quickly determine if the article is really relevant or not.

4.4. The top level—the catalogue

In keeping with our original commitment to an integrated catalogue of scholarly resources, we have chosen to catalogue our electronic journal collection so that access is provided at the highest level along with the body of mainly paper-based scholarly literature available at FIS and throughout the University of Toronto.

Cataloguing standards for Internet resources are in their infancy and continue to be in a state of flux. The basic philosophy of the FIS Library is to integrate these resources into our present catalogue so that our users need look in only one place for items that are most central to our collection. We have therefore decided to catalogue these Internet items according to the same standards used for the rest of the collection.

Chapter 9, Computer Files, of the *Anglo American Cataloguing Rules*, 2nd edition, 1988 revision, *Amendments 1993* (*AACR2* 1988, 1993) is the basis for our bibliographic record description. Through practical application of the present rules, cataloguers have identified some changes/updates needed to deal with Internet resources. Most of these changes are being proposed through the American Library Association Committee on Cataloging: Description and Access, of which the dean of FIS, Lynne Howarth, is a voting member.

Our bibliographic records are stored in MARC format (Library of Congress 1994). The most important new field relating to Internet resources is 856, the Electronic Location and Access Field. The information stored in this field allows for the creation of a hypertext link between the bibliographic record and the actual item. The 856 field carries the URL of the Internet resource. An example of a MARC record with 856 fields is shown in Figure 6.

The final cataloguing standard used is *Library of Congress Subject Headings* (Library of Congress 1995) to provide controlled subject access to the electronic journals in a manner consistent with the rest of our collection.

Information on cataloguing Internet resources is available on a Web page created by Joe Cox, Lynne Howarth, and Trina Richard for their continuing education workshop on the subject.

Our bibliographic records are searchable via the Web interface to the UTL union catalogue, UTCat, which is part of a DRA integrated library system. A search for an electronic journal by title results in a brief display of the catalogue record with hot links coded in 856 fields to the resource itself. Figure 7 shows the brief record for *The PACS Review* with hot links to our Web-based search

```
Record #1  Format: Computer File (Text)
 LEADER _____cms 22_____ a 4500
 001   AVX-4485
 008   900201c19909999txutr1p   d  0  a0eng d
 010   a sn 90000811 $
 022 0  a 1048-6542 $
 035 00 i LCMARC/ACT-8665/COX $
 040    a NSDP $b eng $c NSDP $d OUCA $d NSDP $d CaOTULS $
 082 10 a 025 $2 11 $
 090 0  a   $b FIS Web $
 245 04 a The Public-access computer systems review $h [computer file] $
 246 13 a PACS review $
 250    a [Electronic ed.] $
 256    a Computer data. $
 260    a Houston, TX : $b University Libraries, University of Houston,
        $c 1990- $
 310    a Three times a year $
 362 0  a Vol. 1, no. 1- $
 500    a Title from title screen. $
 516    a Electronic journal. $
 538    a Mode of access: through FIS Web Server:http://        www.fis.utoronto.ca/ejournal/pacsr.html ; through FIS
   gopher:gopher://gopher.fis.utoronto.ca/11/
        library/ejournal/pacsr ; also available through electronic
        mail on BITNET and Internet; Send an e-mail message to:
        (BITNET) LISTSERV@UHUPVM1 or (Internet) LISTSERV@UHUPVM1.UH.EDU
        that says: Subscribe PACS-P First Name Last Name; also available
        through a subscription to the Public-access Computer Systems
        Forum, PACS-L@UHUPVM1.BITNET. $
 540    a Also available in an annual print ed. under the same title. $
 650 0 a Libraries $x Automation $x Periodicals. $
 650 0 a Online catalogs $x Periodicals. $
 710 20 a University of Houston. $b Libraries. $
 775 1  t Public-access computer systems review (Print ed.) $x 1063-164X $
 856 7  u http://www.fis.utoronto.ca/ejournal/pacsr.html $2 http $
 856 7  u gopher://gopher.fis.utoronto.ca/11/library/ejournal/pacsr
        $2 gopher $
 856 0  a uhupvm1.uh.edu $a uhupvm1.bitnet $f pacs-p $h listserv
        $i subscribe $
 856 0  a uhupvm1.uh.edu $a uhupvm1.bitnet $f pacs-l $h listserv
        $i subscribe $
```
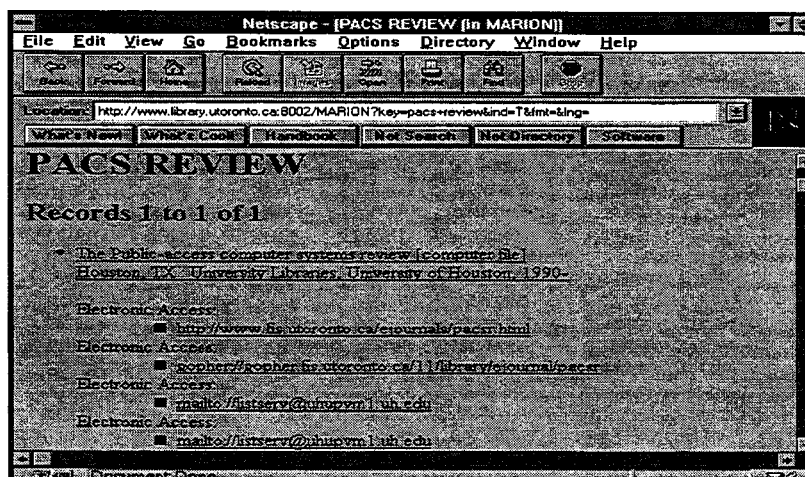
Figure 6. MARC record with 856 fields

Figure 7. A UTCat brief record with hotlinks

form and our holdings for the journal at FIS. Another link could connect to the home page of the journal itself in Houston, Texas.

In utilizing this capability of UTCat, FIS became the first library at the University of Toronto and one of the first in North America to provide immediate access to an Internet resource from an OPAC. This resource at FIS is now easily available to any user of the World Wide Web.

4.5. Unique aspects

There are several unique and innovative aspects to our electronic journal work. First we provide for the search, retrieval, and display at the article level instead of the whole issue. Separating issues into articles results in improved relevancy of hits, since keywords or phrases cannot be in different articles in the same issue file.

If the issue comes as one article per file with another file containing the contents of the issue, the contents are joined to the articles to create one file but maintain article accessibility. This approach reduces the duplicate hits that might result from a WAIS-type search.

The hit list display provides a full citation instead of simply a file name to help in the determination of the usefulness of a hit at an earlier stage. Boolean and proximity searching capabilities are provided. Search terms or phrases are

highlighted and linked within an article to quickly allow determination of the relevance of the article.

It is possible to search an individual journal or the whole collection and always know what is being searched and where the hits come from.

With the addition of bibliographic records to our online catalogue, we have joined the older forms of information with the new for the convenience of our users in an integrated catalogue and also have provided immediate access to the Internet resources from anywhere on the Internet.

5. Future directions

From the above discussion it is obvious that we have concentrated on the collection building and improved access aspects of the project. The whole process of acquiring journal issues, adding them to the collection, and then marking them up to make them searchable is much more labour-intensive than we had expected. We have been unable to incorporate the work into the existing operations of the Library; increased time-saving through automated checking and staff training are needed for this further step to be successful.

The problem of HTML documents still needs to be addressed, particularly as most new electronic journals are now being published in that format rather than flat ASCII. Such journals present problems for a number of different reasons. They already contain SGML-compliant markup that must be retained (but ignored) when the journals are retrieved for indexing. It can be difficult to establish where the borders of a hypertext journal are: What part of the document is to be considered the journal proper, and what part is reference and shouldn't be indexed? These journals also are more subject to change than print or even flat ASCII journals, for editors can easily make corrections or other changes to their official copy of the journal.

We think it would have been useful to design into our DTD a better and more extendable way of labelling the type of article. Currently we support only articles and reviews; having marked up a number of journals, we would like to be able to distinguish among scholarly articles, news items, conference announcements, and others. A DTD design that would allow article types to be easily added, perhaps with SGML attribute tags, would be very useful.

We would also like to develop a Z39.50 server interface for our collection. Currently the University's catalogue can be searched via Z39.50, although users only see a Web-browser form interface. Z39.50 accessiblity would allow the journals' content to be searchable on the same screen and at the same time as the user searched the main online catalogue.

Our project team has been involved with the OISE Library to develop a collection of education electronic journals along the FIS model.. We are also

working to extend this approach to other types of electronic documents such as the School of Graduate Studies calendar at UT and a database of innovative uses of information technology in teaching at UT.

6. Publishing implications

If electronic journals are to become a viable means of reporting scholarly research, several issues need to be addressed by publishers and libraries in order to ensure access at the three different levels of the collection, the contents, and the knowledge context.

At the most basic level, these journals must be archived in order to remain part of the scholarly literature. The archival function for the paper-based literature has been performed by libraries. As the publication medium has been changing to an electronic one, libraries have been slow to take up this function, and publishers, who previously relied on libraries for archiving the paper publications, have often not been aware that they need to perform this function at least in the transition stages. Some issues of journals are already gone. As editors and publishers have begun maintaining Web pages for their journals, more thought may be given to maintaining archival copies of earlier issues as well.

Both a standard delivery method and a standard publishing format like SGML need to be accepted by the publishing industry to facilitate both archiving and content access. Once these are standardized, the procedures have been developed by us and by others such as Eric Lease Morgan for the automation of the storage of these journals and, using our tools, automation of the provision of access to the contents of these journals with little additional human intervention.

At the top level of knowledge context, it is necessary to provide access to these journals within collections. At this time access can be provided by cataloguing these titles and maintaining integrated catalogues of paper and electronic documents. Because the Internet allows scholars to easily access resources physically stored at distant locations, it becomes feasible to rely on fewer collections to serve the scholarly community. As this happens, it becomes essential to provide easy access at an even higher level to the locations of these subject collections.

Notes

[1]This collection is available at: http://www.lib.ncsu.edu/stacks/stacks.html

[2]This collection is available at: gopher://gopher.cic.net:2000/11/e-serials

[3]This search is available at: http://www.yahoo.com/

[4]This search is available at: http://www.opentext.com/omw/f-omw.html

[5]From its inception and throughout its development and implementation, the project has benefited from a collaboration of people from different sections of the Faculty, including Gerry Oxford from Information Technology Services, Diane Henderson and Joe Cox from the Library, our former dean Adele Fasick, and Professor Clare Beghtol. Patrick Gignac, a recent graduate, was our research assistant. As Systems and Acquisitions Librarian, Marte Misiek was the project head.

[6]See http://gort.ucsd.edu/newjour/ for NewJour's web page

[7]For gopher access use: gopher://gopher.fis.utoronto.ca/11/library/ejournal

[8]For web access use: http://www.fis.utoronto.ca/library/libejour.htm

[9]Advice from John Price-Wilkin, who was involved in the Electronic Text Center at the University of Virginia and is now with the University of Michigan Digital Library Project, was invaluable. He has worked on providing access to electronic documents in a Web-based environment. (See Price-Wilkin 1994a, 1994b for descriptions of his work.)

[10]An online searchable version is available at:
http://etext.virginia.edu/TEI.html

[11]Their web page is at: http://www.sq.com/

[12]This form is available at: http://www.fis.utoronto.ca/ejournals/doSearch.html

[13]This page is located at: http://www.fis.utoronto.ca/library/cir

[14]To see a demonstration of this capability,

(1) use your web browser to access http://www.library.utoronto.ca:8002/

(2) choose the search option of **Search by...Author/Title/Subject/Call number**

(3) select **Title** as type of search

(4) search for "pacs review"

The hot links listed under **Electronic Access** are from the 856 fields in the MARC record.

[15]This collection is available at: http://www.oise.on.ca/~aeloise/ejournals/welcome.html

References

Anglo-American cataloguing rules, 2nd Edition, 1988 Revision. 1988. Ottawa: Canadian Library Association. Chapter 9.

Anglo-American cataloguing rules, 2nd Edition, 1988 Revision. Amendments 1993. 1993. Ottawa: Canadian Library Association.

International Organization for Standardization. 1994. *Information and documentation—electronic manuscript preparation and markup*. ISO 12083:1994. Geneva: International Organization for Standardization.

Library of Congress. Cataloging Distribution Service. 1995. *Library of Congress subject headings*. 18th ed. Washington: Library of Congress, Cataloging Distribution Service.

Library of Congress. Network Development and MARC Standards Office. 1994. *US-MARC format for bibliographic data, including guidelines for content designation*. Washington: Cataloging Distribution Services, Library of Congress.

Morgan, Eric Lease. 1995. Description and evaluation of the "Mr. Serials" process: automatically collecting, organizing, archiving, indexing, and disseminating electronic serials. *Serials Review* 21(Winter): 1–12.

Okerson, Ann, ed. 1994. *Directory of electronic journals, newsletters and academic discussion lists*. Washington: Association of Research Libraries.

_____. 1995. *Directory of electronic journals, newsletters and academic discussion lists*. Washington: Associaton of Research Libraries.

OpenText. 1994. *Product information: OpenTextSearch*. Waterloo, Ontario: OpenText Corporation.

Price-Wilkin, John. 1994a. Using the World Wide Web to deliver complex electronic documents: implications for libraries. *The Public-Access Computer Systems Review* 5(3): 5–21.

_____. 1994b. A gateway between the World-Wide Web and PAT: exploiting SGML through the Web." *The Public-Access Computer Systems Review* 5(7): 5–27.

Sperberg-McQueen, C.M., and Lou Burnaud, eds. 1994. *Guidelines for electronic text encoding and interchange*. Chicago, Oxford: TEI P3, Text Encoding Intiative.

University of Toronto, Faculty of Library and Information Science Library. 1994. *Collection development policy*. Toronto: University of Toronto.