Hype(r) about the Web

or Using World Wide Web HyperText to Make M[UMPS] Databases Accessible

by Ed de Moel

One problem that many of us have had to address is the desire to make information generally available without allowing just "anyone" access to the environment where that information is stored. This article describes an attempt to make information available through the InterNet (the World Wide Web) by generating hypertext-structured representations of information that is stored in a M[UMPS] database.

How it All Started

It all started when someone asked me whether I was in possession of a certain article about work with M[UMPS] in the mid-seventies. I knew I had the article, but I couldn't remember in which volume of which newsletter it was published. I had also just moved, and most of my literature was still boxed up. It occurred to me that several friends had set up their lists of literature in machine readable form. After obtaining those lists, I was able to load them into a simple database and scan the various titles. I found that the article was published in a *MUMPS Quarterly* from 1982. I was then able to find the bundle from that year in the moving boxes and make a copy. The whole process took about two weeks. I ended up with a small database of M[UMPS]-related literature and a friend who was happy with the copy of the article.

Of course, having a "small database" never seems to be a stable situation. It is now half a year later, and the small database has become as complete as I've been able to make it. I should now be able to find any article that has been published about M[UMPS] over the past thirty years very quickly. Well... good for me. But how about making this source of information available to the M[UMPS] world at large? I could set up a small database with a simple access interface on one of the servers at the company I work for, but that would only make the information available to those who can log into that machine, and today's firewalls may make that a little difficult, even for people who do work for the same company.

Presenting the Information

I kept looking for possible avenues of access to this database. Handing out floppies proved too impractical: not all recipients had database systems that they could use to order and display the information. So, I started to give people the choice of a floppy with the database, or a printed copy of the various works, sorted by author. This still didn't solve one problem: once you have one of these floppies or print-outs, you have only the version of the database created on that date, and you will have to depend on either yourself or me to remember to provide you with any updates. Clearly an online approach was preferred.

That's when I started to think about using the World Wide Web. The number of people who do not have access to the web is rapidly dwindling. Once someone has a browser or viewer installed, there are no additional fees for accessing specific information, and access is often granted despite the presence of firewalls.

http://world.std.com/~demoel/mdc

I already had some information available on "the web" about the MUMPS Development Committee, so I added a hypertext link to the literature database. Well... actually, there are now three links:

Author index
Keyword index
Publication index

Before we get into any details about how these hypertext links were created, let's have a look at the database. For every entity in the database, there is the following information:

- Zero or more authors:
^DBLIT(0,item,1,name)=“”

- One or more titles:
^DBLIT(0,item,2,i)=title

Zero or more publication references:
^DBLIT(0,item,3,i)=reference
with a couple of cross-references to find entries easier:

- List of “first” authors (for creating the alphabetical list):
^DBLIT(“N”,name,item)=“”

- List of all authors:
^DBLIT(“A”,name,item)=“”

- List of keywords in titles (I considered any word in any title to be a “keyword,” unless it is less than two characters, or a simple word like “the”):
^DBLIT(“K”,keyword,item)=“”

- List of “known” publications and newsletters:
^DBLIT(“P”,publication,item)=“”

Of course, I simplified the structures a little in the above summary, but such entries in databases are probably common for M[UMPS]-based databases.

Hyper Text?

By now, most everyone has probably seen displays that make use of hypertext structures. A typical phenomenon is that text is displayed with some words highlighted. There is a pointer device (text cursor or mouse pointer) that can be moved across the display, and when the pointer is activated (mouse-click or enter-key) while it is over a highlighted area, the display is changed to provide more information about the highlighted entity.

Now what does this have to do with M[UMPS] databases?

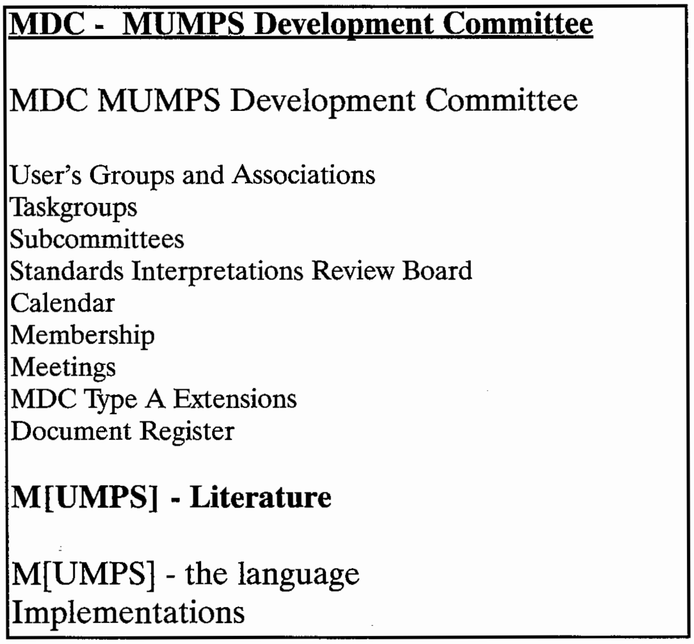
Well, the way we deal with cross-references in a M[UMPS] database is similar to the way hypertext works. Suppose we are looking at the cross-reference on authors. We will see:

```
^DBLIT(“A”, “ackerman,tom”,123)
^DBLIT(“A”, “ackerman,tom”,472)
^DBLIT(“A”, “barnett,octo”,93)
^DBLIT(“A”, “barnett,octo”,127)
^DBLIT(“A”, “barnett,octo”,347)
.
.
.
^DBLIT(“A”, “dvorak,john”,62)
```

(et cetera).

When looking at the list of authors (first subscript is “A”), we can select a specific name, zoom in on the works by that author, then on a specific article, and display the complete information about that article. There is a certain parallel between the way subscripts work in M[UMPS] and the way hypertext links work.

In order to make it possible for a hypertext user to find an article by a specific author, the user must move through a number of “pages” in the hypertext files. When a “webbrowser” displays the information at the location http://world.std.com/~demoel/mdc, the first lines of the MDC “home page” will be displayed.



MDC - MUMPS Development Committee

MDC MUMPS Development Committee

User’s Groups and Associations
Taskgroups
Subcommittees
Standards Interpretations Review Board
Calendar
Membership
Meetings
MDC Type A Extensions
Document Register

M[UMPS] - Literature

M[UMPS] - the language
Implementations

Next, the user moves the cursor to the highlighted text “M[UMPS]-Literature” and either clicks the mouse, or presses the enter-key. When this happens, the browser displays the information that is stored for the highlighted text. In this case, the information is:

<a href=”#mlit”>M[UMPS]- Literature</a>

The text href="#mlit" tells the browser to jump to the section in the current document that starts with the label <a name="mlit">, and the text between > and </a> is the text that will appear as the highlighted visual link to the label.

The information found following that label will cause the browser to display the next page:

M[UMPS] - Literature

Click or press **here** for information about the M[UMPS] literature list.

The literature list can be accessed by **author**, by **keyword in title context** index or, if the article was published by a MUG or an MTA, by **publication,** in which the article appeared.

On this page, the user can select any of the highlighted possibilities (author, keyword or publication), and from there, new pages will appear, and the search path narrows down, step by step, until the desired document is found.

In this case, the codes that allow the browser to branch look like:

```
The literature list can be accessed by
<a href="lit/litauth.html">author</a>, by
<a href= "lit/litkwic.html">keyword in title context</a>
index or, if the article was published by a MUG or an
MTA, by <a href= "lit/litpubl.html">publication </a>, in
which . .
```

A reference like "lit/litauth" causes the browser to open a new file which will display the next page. (In this case the file is called libauth.html and resides in the sub-directory lit relative to the current file.)

In the first version of the hypertext access to this database, the page that appeared when author was selected, contained the names of all authors in alphabetical order. Now that the database has grown to non-trivial proportions, it has become awkward to wait until all (almost 2,000) names have been transmitted. So instead, a page with short-cut branches appears:

Authors of M[UMPS] related Literature

Jump quickly to the names starting with the letter:

a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, y, z

For every letter, the branch code looks like:

```
<a href="litaud.html">d</a>
```

It is probably no longer a surprise that these file names were generated by a program that $ORDERed through the author cross-reference and created a file called "litau" followed by the first letter of the names of the various authors.

And indeed, when the letter d is selected, the names that start with that letter appear:

d

Veronic Dahl, Ola-Olu A. Daini, . . . John C. Dvorak, Miroslov Dvorak, Anthony M. Dymond, Ratislav Dzurik

Veronica Dahl
MUMPS and Prolog: An Unbeatable Combination

Ola-Olu A. Daini
MUMPS in Nigeria

At this point, the user can click on the name of the author, and under the author see the various articles by that author.

John C. Dvorak

Dvorak on M[UMPS], Part I
Dvorak on M[UMPS], Part II

And finally, when we click on the name of the article, a full description of that article displays:

Dvorak on M[UMPS], Part II
John C. Dvorak
M Computing, M Technology Association, Silver Spring, MD, Volume 3, Number 5, (November/December 1995), pp. 16-20, ISSN 1072 3226

And Now for the Big Question

How much is done in M[UMPS] to create these files? Well, not too much. All it takes is a couple of $ORDER loops to scan the database and write information relevant to the subscripts encountered into sequential files in a format that can be “understood” by an HTML (Hypertext Markup Language) browser. HTML is the language of the web.

In this case, the files that were created were:

--- the files that describe the individual articles

- $ORDER through list of “first authors”

- Insert a label for every article, so that a browser may jump to this location and display the information about the article, e.g.:

```
<a name= “Ldvorak001842”></a>
<b>Dvorak on M[UMPS], Part II</b><br>
<i>John C. Dvorak</i><br>
M Computing, M Technology Association,
Silver Spring, MD, Volume 3, number 5,
(November/December 1995), p16-20, ISSN 1072
3226<p>
<hr>
```

The code <a...></a> establishes the label that the browser jumps to when the article is selected. Text between <b> and </b> will appear in bold face. Text between <i> and </i> appears in italics.

- After every 50th file, close the current file and create a new one, so that the amount of text that needs to be transmitted will be limited. We could write all information to one large file, but that would mean that the complete information of that file would be transmitted across the network every time someone wanted to see document information. The current set of files consists of about 110 files with 50 document descriptions in each.

- Create a new cross-reference that will tell us which sequential HTML file displays information about a certain article.

--- Then create the index-files for the authors:

- $ORDER through the list of all authors and create the page that displays all first letters (note that at this moment, we don’t have any author whose name starts with “X” so that letter does not appear in this page).

For every letter that is encountered as the first letter of a name, a line like:

<a href=”litaud.html”>d</a>

is written.

- $ORDER again through the list of all authors, and create the files that link to the documents by the authors whose name starts with a specific letter. At this iteration, just create the header part, where the names of the authors are listed. For every author, a line like:

<a href=”#K_2724”>John C. Dvorak</a>

is written.

- $ORDER again through the list of all authors, but now also $ORDER through the sublist of articles by each author. For each author create a section like:

```
<hr>
<a name= “K_2724”></a><p>
<b>John C. Dvorak</b><br>
```

- The code <hr> displays a horizontal line to create a visual separation.

- The code <a...></a> is the label to which the browser may jump (the value

2724 happens to be the unique internal entry number for the author in the database, but one could use any value that is unique within the file where the label occurs).

- The codes <b> and </b> cause the name of the author to display as bold face.

- The code <p> causes a paragraph break (line end plus vertical white space), the code <br> causes a line break (no subsequent vertical white space).

For each article create a section like:

```
<a> href= “litli010.html#Ldvorak001842”>
Dvorak on M[UMPS], Part II</a><br>
```

- At this point, there is minimal identifying information about the article. But there is a branch to the document, using the label (litli010.html is the file name, and Ldvorak001842 is the label in that file) that was created during the very first step. By branching to that label, the full description of the article may be displayed.

Once again I have used an internal entry number in my database to ensure that the label is adequate and unique (1842 is the internal number for the article in question), but any other unique code could be used for the label in this context.

--- Perform three similar $ORDER loops for the keyword index.

---Perform three similar $ORDER loops for the index by known publications.

After the sequential files have been created, they will need to be moved to a special directory on the web server where they are accessible by anyone in the world who has a web browser program.

Some Considerations and Hints

After about half a year of attempts to format information in HTML, I have learned to avoid several problems:

- It is very easy for files to get “large.” Keep in mind how much information has to be sent across the network to build your information page on the display of the enduser. Experiment with methods to keep the individual files as small as possible. (Those of you who have seen my pages will probably have noticed that I’m still having problems with this rule.)

- The Web is truly world-wide. You may be looking at comments from very unexpected corners of the world about the information that you made available. Try to keep information factual and minimize remarks which may be interpreted as culturally “colored.”

Graphics are great. Indeed, I would like to add more pictures to my “dry” data files. But I have also noticed how long it takes to view pages that other people have built. Sometimes it takes several minutes to download all the graphics files that illuminate a page. When someone is looking for information and you’re keeping them waiting with pictures that do not help, they may quickly turn away from the information you are trying to make available.

- There is a (storage) price to pay. Consider the replication of data that occurs in this example: the (minimal) identification of the various articles is replicated under each author (one article can have multiple authors), under each keyword (I was pretty “generous” in my definition of keyword) and under each publication. In the M[UMPS] database, the same identification only takes as much space as needed to store the internal unique ID codes (numbers in this case).

The M[UMPS] global variable containing the database is just over 2 megabytes in size, the sum-total of the HTML files is almost 4 megabytes.

- Don’t throw away the programs you’ll write to generate these hypertext files. Every time the database has been modified, you will want to re-generate the suite of HTML files.

- Some environments have file names in upper case only, some prefer lower case, some are case sensitive, and some are not. Be consistent in your use of upper and lower case for the names of your hypertext files.

In Conclusion

Overall the result is well worth the effort. For a minor investment, the return is:

- Truly "global" access to data

- In this case: the desired read-only access, so that there are no security concerns related to data integrity. (Be sure to see Jim Self's "Guest Book" for an example that is not read-only and still secure.)

- The interface that suits the end-user: CHUI or GUI, the browser that the end-user employs decides the interface (e.g., Lynx™ is CHUI, Mosaic™ and Netscape™ are GUI, all use and present the same information; all are available on many platforms). *M*

Ed de Moel heads the MDC and works on tools for performance monitoring at Science Applications International Corporation. His e-mail address is: demoel@fwva.saic.com.