



The Personal Virtual Library

[CERN]

by *Federica Brugnolo & Jean-Yves Le Meur*
8/4/98

CERN-OPEN-98-019

Abstract

Looking for "library" in the usual search engines of the World Wide Web gives:

- *"Infoseek found 3,593,126 pages containing the word library" and it nicely proposes: "Search only within these 3,593,126 pages ?"*
- *"Yahoo! Found 1299 categories and 8669 sites for library"*
- *"LycOs: 1-10 von 512354 relevanten Ergebnissen"*
- *"AltaVista: About 14830527 documents match your query" and at the bottom: "Word count: library: 15466897" !*
- *Excite: Top 10 matches and it does not say how many can be browsed...*

"Library" on the World Wide Web is really popular. At least fifteen million pages are supposed to contain this word. Half of them may have disappeared by now but one more hit will be added once the search robots will have indexed this document !

The notion of **Personal Library** is a modest attempt, in a small environment like a library, to give poor users lost in cyber-libraries the opportunity to keep their own private little shelves - virtually. In this paper, we will look at the usual functionalities of library systems before focusing on the description of the Personal Virtual Library (PVL). We will finally explain how it can easily be implemented on the top of any library system.

The need for a personal library

The need for keeping personal data selected within a global information system has always existed. Here are some examples of this need in the world of libraries and Internet.

- **The "dusty" shelves**

Even before the existence of electronic libraries, two separate types of paper libraries existed: public library and private "in-house" library.

Everybody has at home a library: a set of books, maybe some periodicals or documents kept on physical shelves or wherever in boxes, classified in a specific order or randomly, from given publishers and with possibly some notes on them. They are kept because they have been read and appreciated, or they are waiting for being read, or the cover just looks nice.

They are two things that cannot be reproduced virtually, the pleasure of the ownership and the pleasure of the appearance, but all other features of a real private library are normal required features of a modern electronic library system.

- **The Selective Dissemination of Information (SDI)**

Long before the birth of the World Wide Web in the beginning of the 90's, electronic cataloguing and retrieval systems with text only (telnet based) user interface were installed in libraries. These systems included the feature called SDI acting as what is called today *profile services* or *user-centered push system*: it consists in warning users of new incoming documents depending on criterias they have given to the library system.

- **The bookmark system**

With the very first browser of the World Wide Web, the option of keeping bookmarks was already available. The bookmark system is actually a client-side personal library of URLs. The need of keeping a private list of links by Web users is obvious with the amazing multiplication of "personal pages".

- **The on-line supermarkets**

The possibility of marking an element in a set of many elements first appeared on the Web with on-line supermarkets offering the electronic shopping opportunity, often called *electronic-basket*. It acts as a server-side personal library where the choice of a user is memorized by the server till the user gives his credit card number. The user selection is only kept for a session and it is lost when the user logout the browser.

The **Personal Virtual Library** is one step beyond the on-line supermarket as it intends to keep marked documents through multiple sessions and to offer many options on these documents.

Overview of the minimum required functionalities

We list in this part the basic features that a user of an electronic library on the Web should expect, independently of any personal interaction.

- **Latest documents**

The most recent documents arrived in a library must be very easily available to the readers. This is particularly true for the catalogues of research papers where users are keen to overlook at the very last publications.

- **Text only and Java versions**

Both must be provided. The distortion between what the different browsers can support makes almost impossible to provide a technologically advanced graphic interface working for the latest versions of the browsers. The possibility to quickly switch to a text-only version is therefore a safe solution for ensuring a wide access to the library. This is particularly true at CERN where some physicists are still using obsolete browsers (like *lynx* which only displays text).

- **Navigation through catalogues**

All libraries keep different types of documents which can be browsed separately. Hypertext navigation through the various catalogues, and possibly through categories, dates or other criterias is something expected by all readers.

- **Flexible searching scope**

Before starting a query, it is important that the scope in which the request is run can be reduced to a specific sub-set of the library database or enlarged to multiple catalogues, and possibly to multiple libraries if an inter-library protocol (usually Z39.50) is enabled.

- **Simple search**

By analyzing the log files of the http server, it has been stated that most users are running elementary queries. The ordinary behavior is to type a word and to select the field for searching (author, title, etc.). This option must be straightforward from all pages of the library site.

- **Refine a search**

When a large number of results is found, the users expect to be able to detail what they look for. Giving the opportunity to refine a search (searching within a list of documents) is a good way to simply build a complex query.

- **Complex search**

A complex search may include a combination of logical operators (AND, AND NOT, OR) between different texts and search fields. It may require the results to be sorted according to a given code and it may run across many catalogues, subjects or periods.

- **Ask for next records**

The user can move directly to any position in the list of records and he can display at once as many records as he wants.

- **Show more information**

The list of the records gives a first short view of all documents. A detailed view of each matched result can be obtained by a click and it provides a full bibliographic description.

- **Access to the document**

If the document is electronically available (on the Web), a link should be proposed to read it on line (or get it and print it). If it is a published article, some special rights might be needed to download the full text of the document.

If not, readers should be proposed to look at the holding status and to eventually ask for a loan. The inter-library loan can also be available (depending on library systems and agreements).

- **Browse a field**

Another way to find a document is to scan (or browse) the indexes of the database by alphabetical order. The reader provides the starting word and he can see other words appearing in the database. This is particularly useful when looking for report numbers.

The personal options

The Personal Virtual Library is made of multiple options. Any user is identified when he is browsing the catalogues and he can modified his own view of the library. Before entering into the technical description of the personal library, we describe here the options available.

- **Create a personal shelf**

The first step in creating a Personal Virtual Library on top of an electronic library system is to enable the option of "marking" records. This is now an usual option of electronic supermarkets but it is still rare in the library world. In fact, during the navigation and the search into a Library Catalogue, the user may be interested in several documents fetched in different pages and requests. The aim of the *personal shelves* is to allow the user to mark these documents, to retrieve them in a personal page and to provide further functions of management of the data.

The functions proposed are:

- **see the *personal shelf* page:** this page contains all the documents marked by his owner;

- **mark/unmark documents:** the user can mark/unmark a document from a generic output search page, from the *personal shelf* page or from the "full format" view of a document;
- **reload shelf:** this function allows the user to reload his *personal shelf* page. To be used, for example, if he defines a *personal format*;
- **add shelves:** this option will merge two different shelves. All the marked records in the first shelf will be added to the second one.
- **clean shelf:** it will remove all the documents stored in a *personal shelf*.

To have more details, refer to [1].

● View personal searches

The user could be interested not only in saving documents, but also in particular searches. This service lets the user sum up the searches he runs during a session and eventually save them in his PVL, giving them a *nickname* (see fig. 3). For each request stored the user can access the result by a simple click from the *personal searches* page.

The service is divided into two parts. The first one is a list of the searches the user has performed during the current session (see fig. 1). The second part is a summary of the queries kept in the PVL (see fig. 2), this part proposes various actions:

- **delete a query:** to remove a query from the current PVL;
- **rename a query:** to change the nickname of a query;
- **combine two query:** to combine two queries. It actually runs the first query AND the second query.

Few examples of screen representation are shown in the diagrams below. To have more information consult the [2].

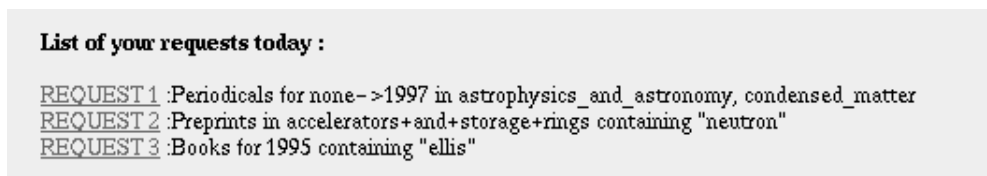


Fig. 1: List of the requests of the current session

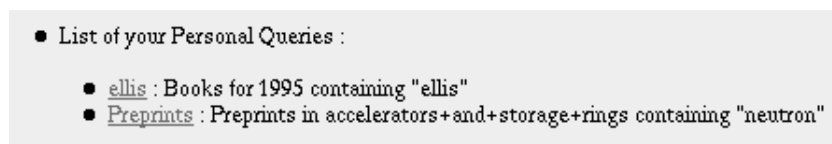


Fig. 2: Summary of the query kept in a PVL

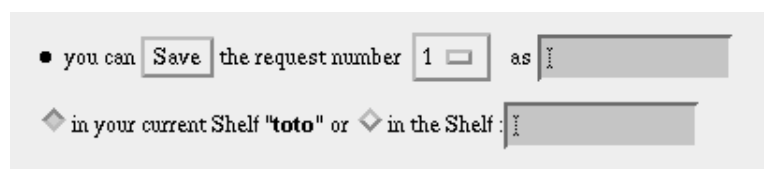


Fig. 3: Save option

- **Set up personal formats**

The owner of a PVL can define his own output format for displaying the result of a search by selecting the information he wants to display. If he does not define any *personal format*, a standard one is used.

There are two ways to define a Format Output:

1. **the Table Description Format** (see Fig. 4): the user can select some checkbox buttons, with special meanings:
 - **Show**: it shows the associated field
 - **New Line**: it inserts a new line after the field value;
 - **Bold**: it shows the field in bold
 - **Italic**: it shows the field in italic
 - **Position**: it defines the position of the field
2. **The Dynamic Markup Language Format** (see Fig. 5): the user can digit the HTML code that will be used to write the output search pages. Besides the usual HTML tags he can write the fields using their \$\$CODES. If the code inside a line does not exist for a record then the whole line is not displayed.

Show	New Line	Bold	Italic	Position
<input type="checkbox"/> RN - Report Number	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1
<input type="checkbox"/> ER - Server Number	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2
<input type="checkbox"/> TI - Title	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	3
<input type="checkbox"/> AU - Authors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4

Fig. 4: Table Description Format

```

<font color=red>$$YR</font>
, <code>$$RN</code>
<b>$$TI</b>
, <i>$$AU</i>
$$MINFO
  
```

Fig. 5: Dynamic Markup Language Format

For further details consult the [3].

- **Set up personal profiles**

A *profile query* is a set of rules that documents must satisfy in order for them to be considered interesting. The *profile query* is a search which can be regularly executed with the results notified to the Query owner. Using the PVL the query owner could also define a *personal shelf* where the new documents could be stored, using a *personal format* if any.

The parameters of a *personal profile* are:

- the **nickname** of the profile, chosen by the user
- the user **email address**
- the **frequency** of notification
- the option of feeding a **personal shelf**
- the option of receiving an **alert email**

The actions available are:

- **list** all the profiles stored by a user
- **remove** a profile
- **remove all** the profiles in the PVL.

- **Insert personal annotations**

Consulting a library catalogue, containing a big amount of data, the user could have the feeling to be lost in the output results. A solution to avoid this problem is to allow the user to write annotations and in this way, to keep and possibly share important information related to the documents.

The *personal annotations* can be of different types:

- texts
- images
- generic files
- URLs

They are displayed next to the document, but only the owner of the PVL can see them.

- **General options for all the services**

There are a set of functions which allow the user to consult and modify his PVL:

- **Change Identification**

The user can change the PVL identification, like he can change a password. Each Personal Virtual Library is identified in a unique way by the Id chosen by the user [see next chapter].

- **Show other shelf**

A user can define an indefinite number of PVLs. For example he can keep a private shelf for his personal interests and he can share another PVL with his group of work.

- **Merge**

Using the merge functionality two different PVLs could be unified in one.

The Identification systems and the Personal Library structure

The first main technical issue for setting up personal options is to decide how users are going to be identified.

● Identify the reader

After a reader has selected records, there are two different views on what to do with these:

1. keep them in memory for the time of the session: it is a session-only markup. The objective can be to gather products - books - for borrowing them or - preprints - for printing them at once, etc.
2. keep them in an allocated space (called e-shelf) for future reference. The selected information remains through sessions and can be accessed at any time by any browser.

It is clear that these two types of markups answer to two distinct needs, the immediate consummation or the longer term preservation. The first need is shared by sellers and libraries while the second one is obviously very library related. Knowing from the beginning the need you are facing is essential to determine the computing technique you will use for the markup system.

1. "One-session" identification

The most appropriate technique is certainly the use of *persistent client state http cookies* (see [5]).

As a personal library aims at keeping personal information in a long term, this technique is not satisfactory and we will not enter into the details of its implementation.

2. "Long term" identification

The technique for long term identification of the readers depends on the expected level of confidentiality for the personal libraries. Two different methods can be developed.

■ *Username/password* identification.

Asking for a username and password to enter into a personal library or to perform any action related to a personal library is the best way to ensure the confidentiality. It has nevertheless many drawbacks:

- It is heavy for the users. They do not necessarily use the library system very often and they do not want to remember a username and a password for consulting their records. Moreover, users are all over the world, which prevent of using a local existing password file.

- It is heavy for the support team. Accounts have to be created and maintained. It would really generate a major task of user administration which may not be desirable.

- Finally, the need of confidentiality for a personal library built on top of a public library is low. A "light" solution is certainly preferred, even if a high level of privacy is lost.

■ *Hidden nickname* identification.

Each time a user access any page of the library Web site, a unique number (based on the date and the process number of the CGI program creating the page) is attached to the user. When moving within the library catalogues, running searches or displaying documents, this identifier is kept as an *hidden* HTML variable.

At any time, the user can mark a record or define a format, or perform any personal action. He is identified by his hidden nickname, which is by default the unique number calculated at the first access but which can be changed to any name the user chooses.

No maintenance at all is necessary with this technique. Two automatic cleaning procedures run to avoid useless data storage. Every night, the one-session shelves created the day before are canceled. Once a year, a check is run on all personal libraries to remove the ones which have not been looked at within a year. All other actions (renaming a library, merging two libraries, etc.) can all be performed by the readers themselves.

The most performing system would leave to the user the choice of the level of confidentiality, with different access rights for the different actions, using one of the techniques described above. This may sound nice but it may as well make the design much heavier and we did not think it was appropriate for CERN library users.

● **The Personal Library architecture**

The personal library system is actually based on three databases:

1. The main one, the Document Database, contains all information (called meta-data) related to a document, including a unique system number and one or many URL's.
2. The Object Database keeps all full text of the documents, with a unique document id.
3. The Personal Database records all the data related to the users personal libraries. It contains the unique user identifier and the personal information, plus the system numbers of the documents kept in the Document Database.

Below is the diagram summarizing the architecture of the personal actions, from the reader to the databases.

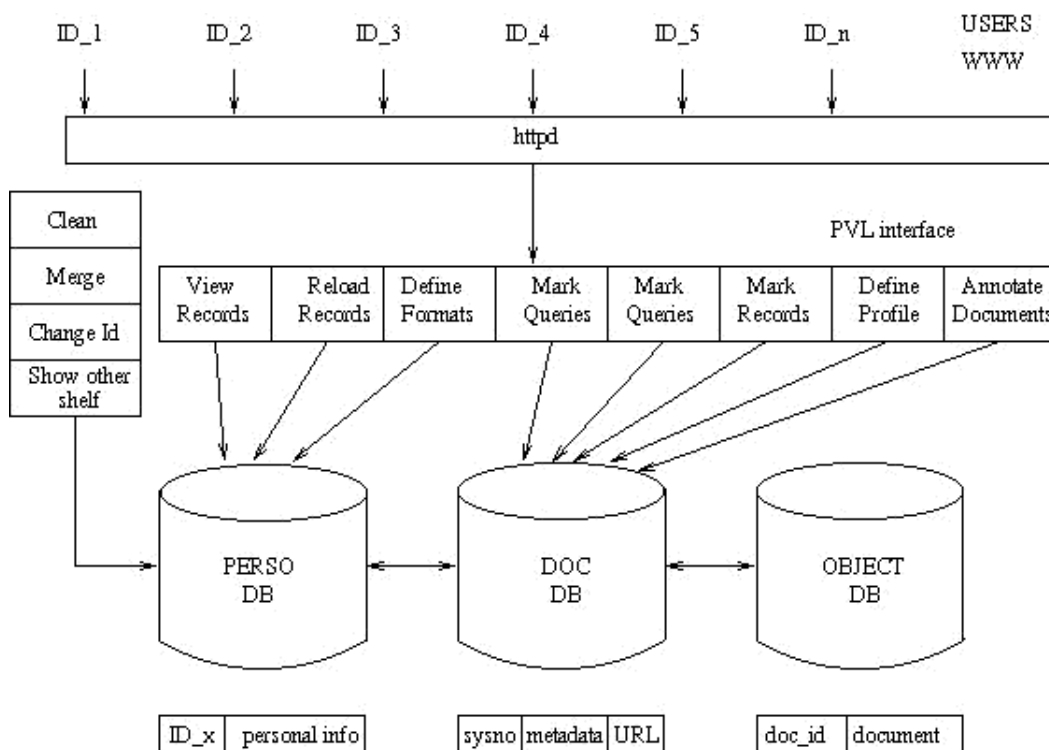


Fig. 6: Identification system and structure

Some actions (like the global ones: Clean, Merge, etc.) directly interact with the Personal Database while others (like marking a query, defining a profile, etc.) need first a request to the Document Database before writing into the Personal Database.

The basic architecture is simple and the combination of the three databases is easy thanks to the

three identifiers: the record id, the PVL owner nickname and the fulltext document id.

Statistics of CERN Personal Virtual Library

Even if the need for personal libraries seems obvious to us (computing support for CERN library), it was important to check the use of the personal tools after the first version was delivered. It gave us the confirmation to carry on with the Personal Virtual Library project.

The first version of the CERN PVL [4] was put into production the 1st of July 1997. Three options were proposed: the personal basket, the personal formats and the personal searches. Here are the statistics of the usage of these services, nine months later.

Personal Shelves	Personal Formats	Personal Searches
788	70	189
88 per month	8 per month	21 per month

Tab. 1: Statistics on CERN PVL: Web-Lib

With almost three shelves created every day during nine months, it clearly shows that the basic feature of the PVL is used a lot. In a few months, we will be able to calculate significant statistics concerning non-used shelves (we give one year delay to our users to access their libraries before removing them).

If keeping records seems to be popular, the use of personal formats is not so frequent. Eight formats are defined each month. It means that the default format is satisfactory to most of the readers. The demand for creating a bibliography or for displaying the information with a personal display is not high. It must be noted that the option is not as straightforward as the "Mark record" option which appears next to all records. The second remark is that the use of the "Dynamic Markup Language Format" allows us to satisfy all the requests concerning the display of the results. It is an extremely useful tool on the support point of view.

Finally, saving a personal search has been used about once every open day. It is much more than we expected. The method for saving a search is not direct (sum up all searches before saving the selected one) and the search system is easy to use. Even so, our users are interested in this "shortcut" option, allowing a fast retrieval on pre-defined criterias.

Globally, these first results are encouraging but we are aware that the curiosity of our users might affect the quality of the statistics. Moreover, the progression should decrease once the usual users have set up their Personal Libraries.

Conclusion: what next ?

The next short term evolution is the addition of the profile service (push technology) and later, the record annotation option. With these two additional features, the CERN PVL project will provide quite a complete kit for personal management of all CERN library catalogues.

In a longer term, the PVL should become an interface layout to be added on the top of many catalogue systems. Whatever database is used, whatever search engine is run and whatever is the formatting of the results, all the options of the PVL can be available. By extension, searching through multiple libraries (with Z39.50 protocol) can also be interfaced using a PVL. Only the notion of the "source library" of a record would need to be implemented.

One of the difficult point in managing a PVL is to propose an "as simple as possible" graphical user interface, yet complete. A study need to be carried out to analyze different approaches for guiding the readers.

Eventually, the cost of a PVL is low. The disk space is not a problem: the information which is kept consists in a small volume (mainly identifiers and short strings are stored). The maintenance is reduced to its minimum by letting users managing their own library and by setting up automatic cleanup procedures.

With a low cost and attractive features, the installation of a Personal Virtual Library is a solution to enable the creation of virtual shelves containing virtual collections of personal data.

References

[1] CERN Library Catalogue: Personal Shelves Guide,

http://alice.cern.ch/cgi-bin/guide_shelf?uid=0

[2] Guide for the Summary of your Searches, http://alice.cern.ch/cgi-bin/guide_viewsearches?uid=0

[3] Report Generator Configuration Guide, <http://alice.cern.ch/cgi-bin/outputformat?uid=0>

[4] CERN Library Catalogues, <http://alice.cern.ch>

[5] http://www.netscape.com/newsref/std/cookie_spec.html
