

The Virtual Notebook System: Distributed Hypertext for Collaborative Research

*Frank M. Shipman, III, R. Jesse Chaney, G. Anthony Gorry
Baylor College of Medicine, Houston, Texas*

Introduction

We are developing the Virtual Notebook System (VNS), a prototype of an information integration and management environment for collaborative work. Our main goal is enhancing the productivity of scientific groups engaged in basic and clinical research in an academic medical center. As the name implies, the VNS is an electronic analog to the scientist's notebook, and it functions as the repository of data, hypotheses and notes, patient information and the like. But unlike the traditional notebook, the VNS is expressly designed to enhance information sharing among the members of scientific teams. A hypertext program we have developed is the foundation for this sharing. Further, the hypertext enables us to integrate into the VNS a variety of computer-based information resources that are important in biomedicine.

Early prototypes of the Virtual Notebook System used Xerox's NoteCards* and later Knowledge System's KMS.*[1] Both of these hypertext systems lack features that we feel are needed in our environment. NoteCards did not allow groups of users concurrent access to the hypertext network. Concurrent access is essential to our support of collaborative work. While KMS provided this capability, it was limited in its ability to incorporate external information resources and applications. Also, we wanted to distribute the VNS over a range of existing hardware and software platforms, many of which NoteCards and KMS did not run on. While other hypertext systems

were discussed in literature, some with quite appealing functionality, we found all lacking for our application.[2] In some cases, the systems were simply not available in a supported form; in others, foreign applications could not be integrated with the hypertext; and for some, highly-specialized hardware or operating systems were required. Therefore, we somewhat reluctantly undertook the development of our own hypertext system for the VNS.

Our principal goals were to provide strong support for collaborative work and facilitate the distribution of the VNS in a heterogeneous computing environment. The result of our ef-

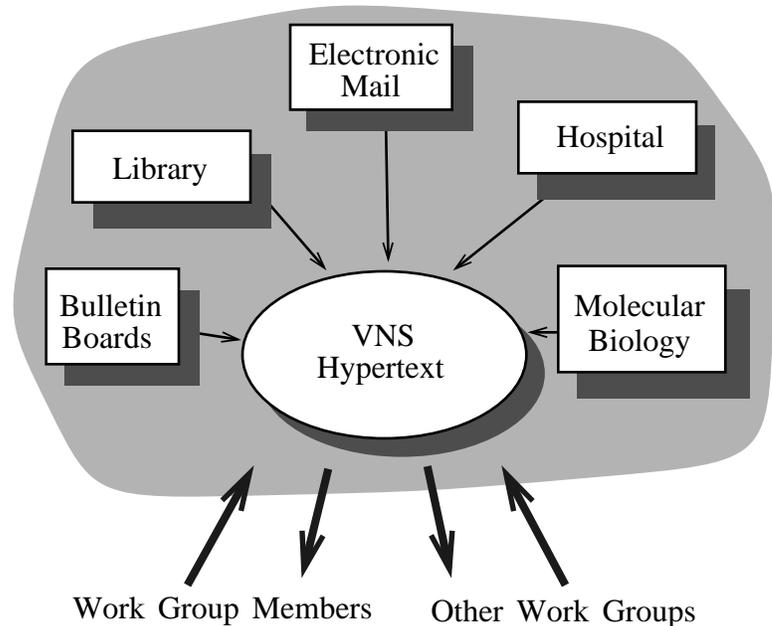


Figure 1. - Information Flow within the VNS

fort, which is described here, is a hypertext system implemented in a relational database (Sybase*). We have maintained a clear distinction between the user interface (X Window System*) and the data access mechanism for the hypertext. This has helped us integrate information resources into the VNS.

While our main goal has been to support biomedical research, we believe that many of the information acquisition and management tools developed for researchers will prove valuable in other settings. Certainly the technologic infrastructure we are developing - the communications architecture and the linkages to information resources and other institutions - will serve many purposes in addition to those of research. Administrative and managerial duties require similar task coordination and integration.

Hypertext for the VNS

The design of the VNS envisions a number of work group servers (WGSs) connected through a network. A WGS would provide local information storage and management by maintaining the group's hypertext in a relational database on the group's server. Users on a given WGS could share information with other groups and access information stored on other WGSs through the hypertext. Integration of the WGSs is an important function of the VNS.

The VNS is to serve as a notebook, so the basic unit in the hypertext is appropriately a **page** which defaults to be approximately the size of a sheet of paper. The user can resize pages within the limits of the screen. Many pages can be displayed on the screen at one time. Pages may be of several types. Some do not allow modification of their contents, as the pages in a researcher's notebook may not be changed. Others allow the addition or removal of information at any time.

A page contains a **title** and any number of **information objects**: text (research notes, electronic mail) or images (graphs, pictures). The objects may be positioned anywhere on a page, even beyond the current visible extent of the page. Information objects can be created either directly by the user or by applications as discussed below. The user can create text objects with an EMACS-like editor or make use of the X Window System's selection service for transferring text from other windows. A "snapshot tool" permits the creation of image objects. The snapshot tool allows a user to sweep out portions of the screen with the

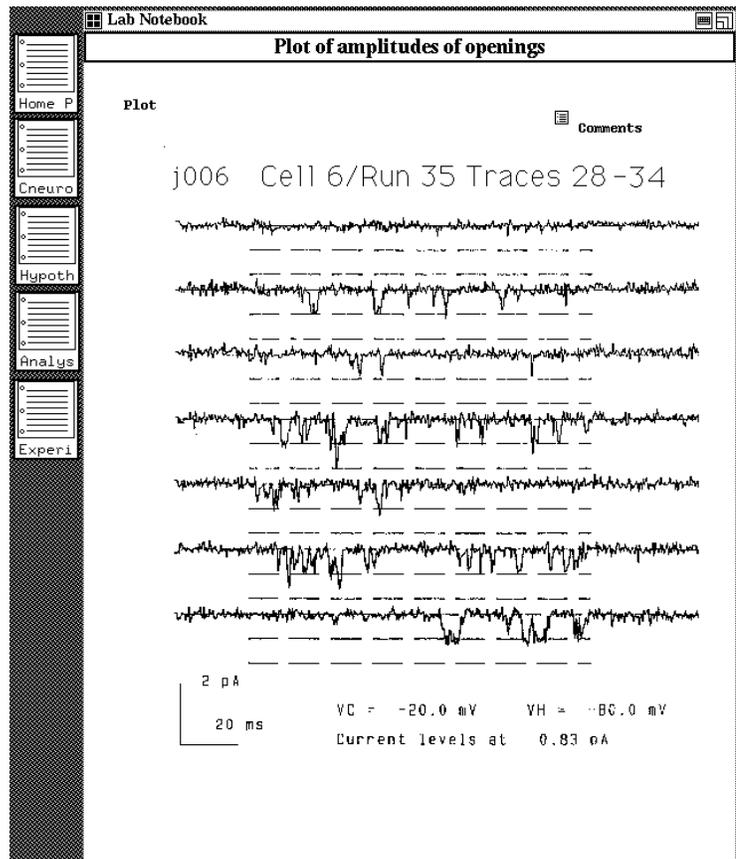


Figure 2. - A Page from the Notebook

mouse cursor and place the captured image into a notebook page. Anything displayed on the screen can be imported into the notebook using the snapshot tool. As discussed below, applications can add information objects using a "paste" facility which positions text or images on a user's page. Figure 2 shows a notebook page with text and image objects and other pages collapsed to an iconic form.

The user can also **copy** objects from existing pages. When objects are copied from pages only a reference to the object is placed in the new page; the object itself occurs only once in the database and is shared between the pages. Indeed, much of the support for collaboration in the Virtual Notebook is based on the sharing of objects through our distributed hypertext system. The information objects which comprise a page may in fact be distributed among several work group file servers.

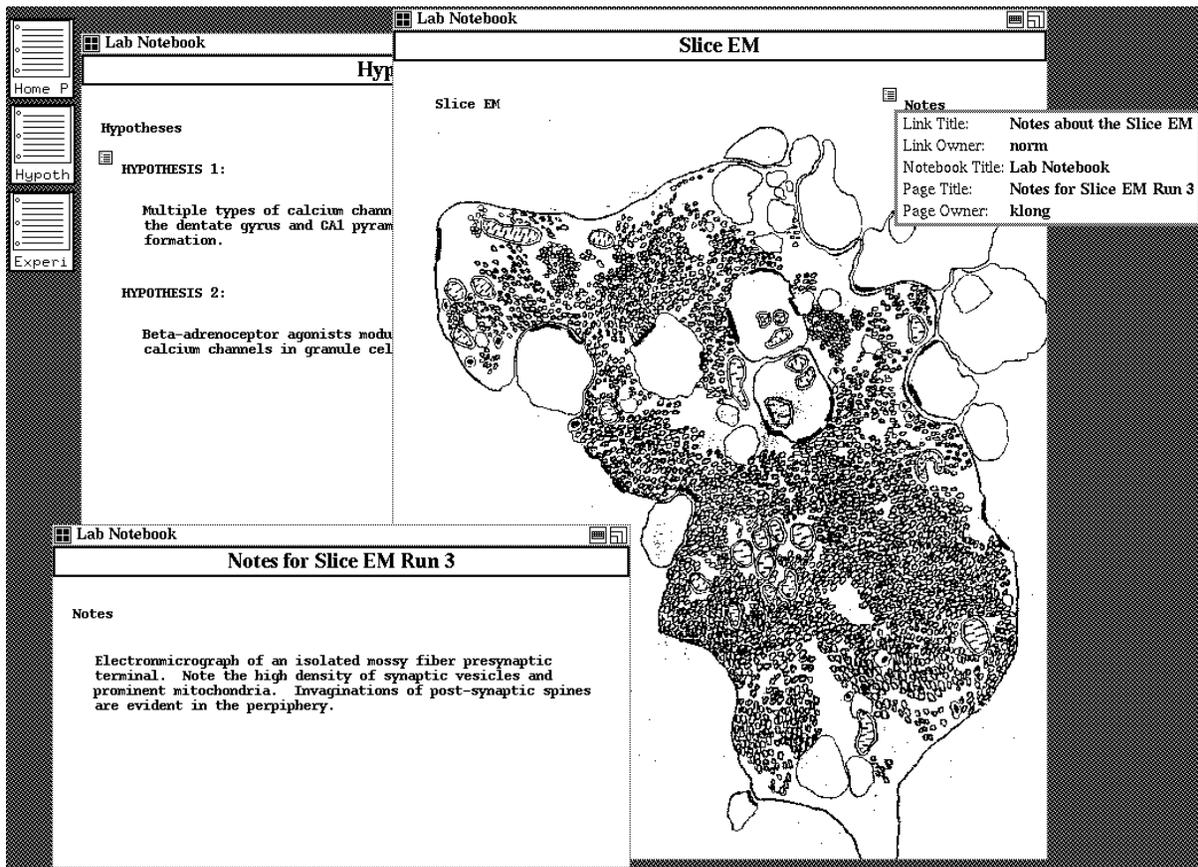


Figure 3. - Notebook Pages and Link Information.

A user can create a new page at any time, adding and editing text, sweeping out images off the screen, or copying objects from existing pages. After the layout and content of a new page are to the user's liking, the page composition (information objects on the page, their location and size), page creator, the page's title, and time of page creation are stored in the relational database.

Pages may contain **links**, references to other pages in the hypertext web. A link appears as a small icon on a page and allows rapid movement from one page to another. When a user moves the mouse cursor over a link a small window pops open providing information about the link and the page to which the link points: link title, link owner, destination page ti-

tle, and page owner. Moving the cursor out of the link closes the information window. This provides users with a way to "preview" the page which the link references. Clicking the left mouse button while in a link causes the linked page to appear in a new window. Links belong to the owner(s) of a notebook and although they appear to be on the page, they are, in fact, stored separately from the page. Two users might share a page, seeing the same information objects on it, but view different links. This is because the page is assembled dynamically for a given user. From the page composition and link information (creator, types of links, and positions on the page), all of which are retrieved from the relational database, the hypertext application assembles a page on demand and overlays it with the appropriate links. Help pages may have links from every user in the system, though each user wants to see only his personal annotations. Figure 3 shows several open pages and window displaying information about a link.

To create a new link on a page, the user simply positions the link in the "source" page and then clicks the mouse in the "destination" page. Information for the new link is stored immediately in the database and the new link appears in the source page.

A **notebook** is a web of pages connected by links. A user may work with many notebooks, some of which may be shared by members of his immediate work group or by other users of the

VNS while some notebooks may be private. A link on a page may point to any other page, so a single notebook may share pages with other notebooks.

A first time user of the Virtual Notebook System is given a notebook with a home page containing a link to his **scrapbook**. The Scrapbook is, as its name implies, a temporary repository for new objects that the user has not had a chance to organize. It is particularly useful in capturing information during interactions with information resources. New pages of the scrapbook are created as needed and each may contain many text or image objects. Each user's home page contains a link to the most recently created page of that user's Scrapbook. The pages of the Scrapbook are automatically linked together in a linear matter from the most recently created page to the oldest page. New objects pasted into the Scrapbook are always put on the most recently created page.

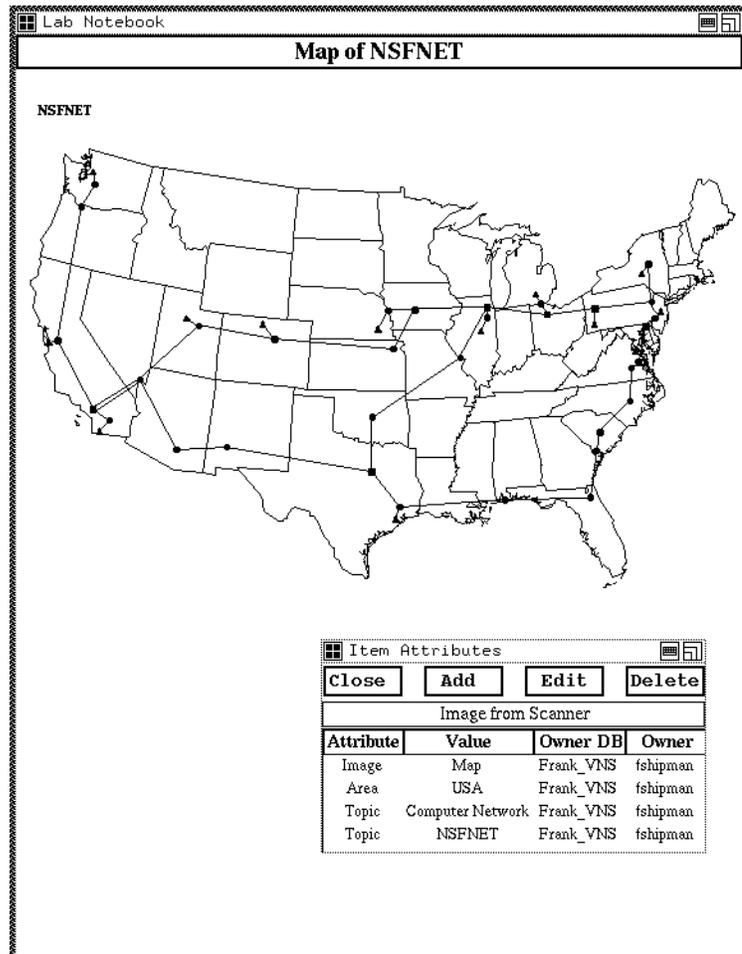


Figure 4. - A Notebook Page with Attributes

Each of the items (information objects, links, pages, notebooks, and users) in the VNS may have attribute/value tuples. This allows additional information to be associated with elements in the hypertext, either by a user or other programs creating information objects and pages in the database. Any number of attribute/value tuples may be added to a hypertext item (see figure 4). An attribute may be multi-valued. Each attribute/value tuple, like a link, is associated with the user who added it, so only attributes created by a specific user could be retrieved.

A VNS mail tool permits the user to mail and receive pages of a notebook. As not all scientists will have access to the VNS, we encode the pages for mailing in PostScript. This allows the page to be displayed or printed on standard PostScript devices. A VNS user may send a hypertext page to anyone who has an electronic mail address and the ability to view or print PostScript.

Database Manager

All of the data in the Virtual Notebook System are stored in a relational database on a WGS. When a page from the notebook is to be displayed, the hypertext program retrieves all the information objects and links and creates a window in which it displays the page. The hypertext has been designed such that when a page contains references to information objects not stored on a user's WGS, those references are resolved through a mapping the VNS maintains of object names and WGSs. The non-local objects can then be retrieved for use in the display.

We have chosen Sybase as our dataserver for its performance and advanced features. We plan to make extensive use of the ability of a Sybase dataserver to make remote procedure calls to other Sybase dataservers on different machines. This will allow us to have a dataserver running on one WGS computer make calls across our network to another WGS at a geographically distant site.

We anticipate that though users of VNS may share data across several machines, most of their work will be done on their local work group machine. This arrangement distributes the processing load and disk utilization in accordance with the group activities. Within the constraints imposed by Sybase, each work group may choose its computer hardware based upon projected computing power, disk requirements, availability of other software, and cost. The decentralization of the hypertext also provides a greater degree of fault tolerance. Machines may be off-line and not prevent everyone from using VNS.

The power supplied by using a distributed relational database makes many other issues for a hypertext system easier to solve. We rely heavily on the Structured Query Language (SQL) calls for its searching power and flexibility. Other programs have easily been developed which make use of these database calls. Problems which would otherwise come up with concurrent access are solved more easily by the dataserver's maintenance of data consistency. Locking mechanisms are provided by the dataserver. Access control, data security, can be implemented at the database level, in addition to the operating system level. The transaction logging provides additional data integrity. Sybase also supports disk mirroring.

Window System

Our program development has been done in the X Window System. The hypertext program runs as a client to an X server. The X server and the VNS may be running on different

computers connected by our network. By exploiting the client/server model of X we are able to distribute computing power and load. Figure 6 shows how X client programs can be distributed through a network.

One of the most important reasons we have developed in X is its potential to become a *de facto* standard for many hardware manufacturers and software developers. We can run the hypertext on several types of hardware: Sun Microsystems,* IBM,* HP,* and Digital Equipment.* X servers are being developed which will allow us to run the Virtual Notebook on MS DOS* compatible computers and Macintoshes.* So scientists can use existing hardware and software, in which they may have a great deal invested, and utilize the functions of the VNS. Furthermore, the development of X Terminals will add to the ease of distributing VNS. They are less expensive and easier to administer than full-featured workstations. An X server running on a PC, a Macintosh, or an X Terminal does have some limitations, however, we feel that a typical user could be well served by such. As requirements rise, a more powerful workstation can be obtained without loss or change of function.

Information Resources

The Virtual Notebook System includes information resources distributed across hardware and software platforms. For standard UNIX* resources like electronic mail, the C shell, and Usenet bulletin boards we have created interfaces that can directly put information into the notebook. This is important to both the computer knowledgeable and the computer novice. For the computer professional our tools will allow the inclusion of the daily routine of reading and responding to new mail, and reading bulletin boards within the framework of the VNS. With one button click, information from the resource is pasted to the user's Scrapbook.

All of the hypertext system's interfaces are brought up through the **Gatekeeper**. The Gatekeeper manages the processes and provides for a seamless interface between the wide variety of programs included in the VNS. This is important to the computer novice who would likely be confused by the need to move from program to program at the operating sys-

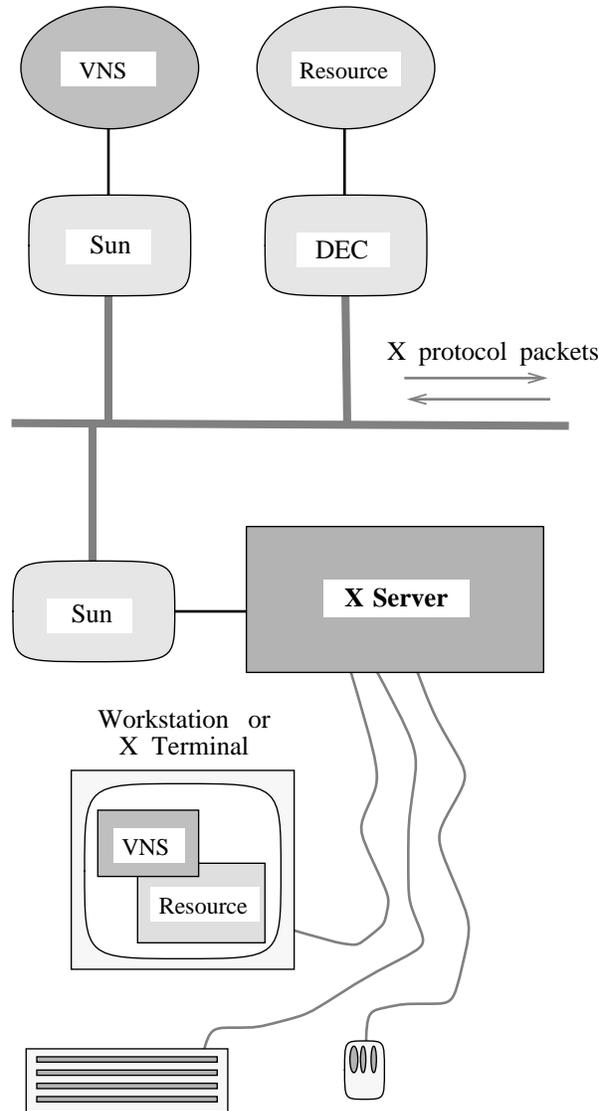


Figure 5. - The Client/Server Architecture of the X Window System

tem level. All of our information resources are mouse and graphics oriented and were written in the X Window System.

While there are many information resources used in biomedical research, a few of obvious importance have already been integrated into the VNS. An important resource for medical researchers is the MedLine database of bibliographic and abstract information on articles from medical journals. MedLine is distributed by the National Library of Medicine. The actual database that we use is the TexSearch database, maintained at the M. D. Anderson Cancer Research Center, in the Texas Medical Center. The interface provided within the VNS connects automatically to make the location and hardware/software details transparent to the user. The TexSearch interface operates similarly to GRATEFUL MED, an MS DOS based product developed by the National Library of Medicine. Our interface to TexSearch is shown in figure 6.

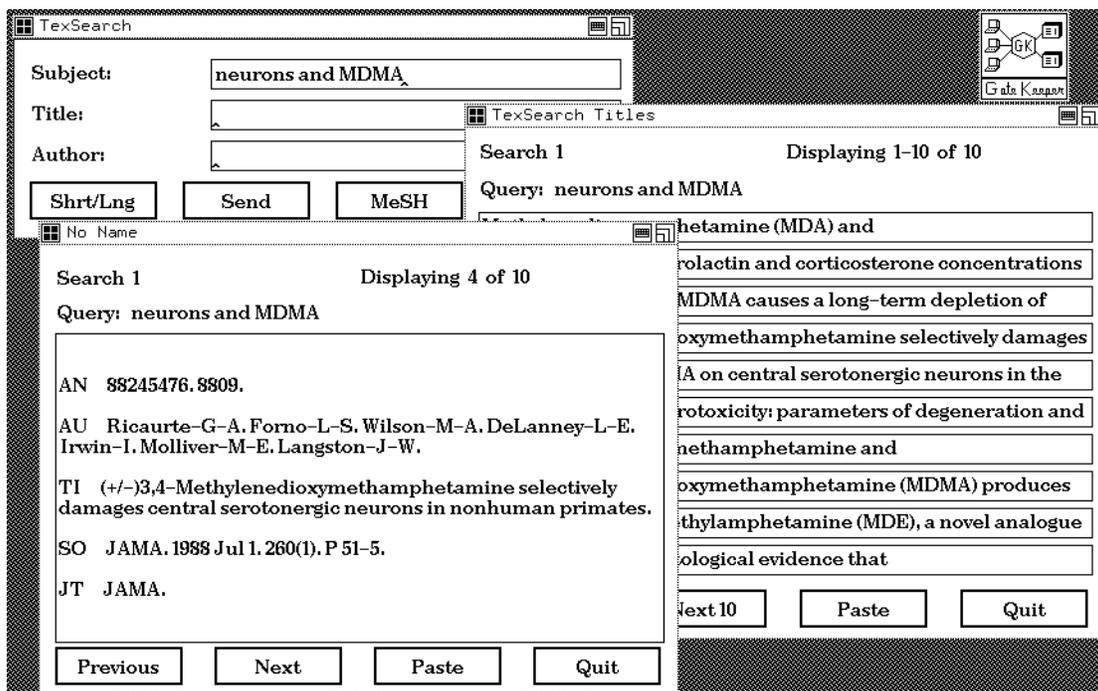


Figure 6. - TexSearch Interface and the Gatekeeper Icon

The Methodist Hospital Information System (HIS) is very important to the clinical care groups we are planning to support, as it contains much of their patient data. Again the interface is completely mouse driven and location transparent for the user. EuGene, a program being developed by the Molecular Biology Information Resource (MBIR) research group here at Baylor, performs nucleic acid and protein analyses commonly implemented by molecular biologists. Besides pasting information into the hypertext, some of our information resources, like the specialized database interface shown in Figure 7, are capable of bringing up pages from the hypertext system. When these pages are displayed by an information resource they may have all of the functionality they would normally have in the hypertext system.

We plan to develop interfaces to other information resources. These will include more specialized database interfaces for the research groups that will be the first groups, outside of our own, to use the VNS. A particularly general interface that is planned is a scanner interface that will allow the scanning of printed documents and images directly into the hyper-

text for annotation. Also, a graphics package for plotting data, as well as a statistical package like SAS,* will be integrated to provide the VNS with the needed ability of viewing and calculating experimental data. Currently Baylor is developing the Baylor Departmental Information System (DIS) and when completed this database of grant and personnel data will be available through a VNS interface.

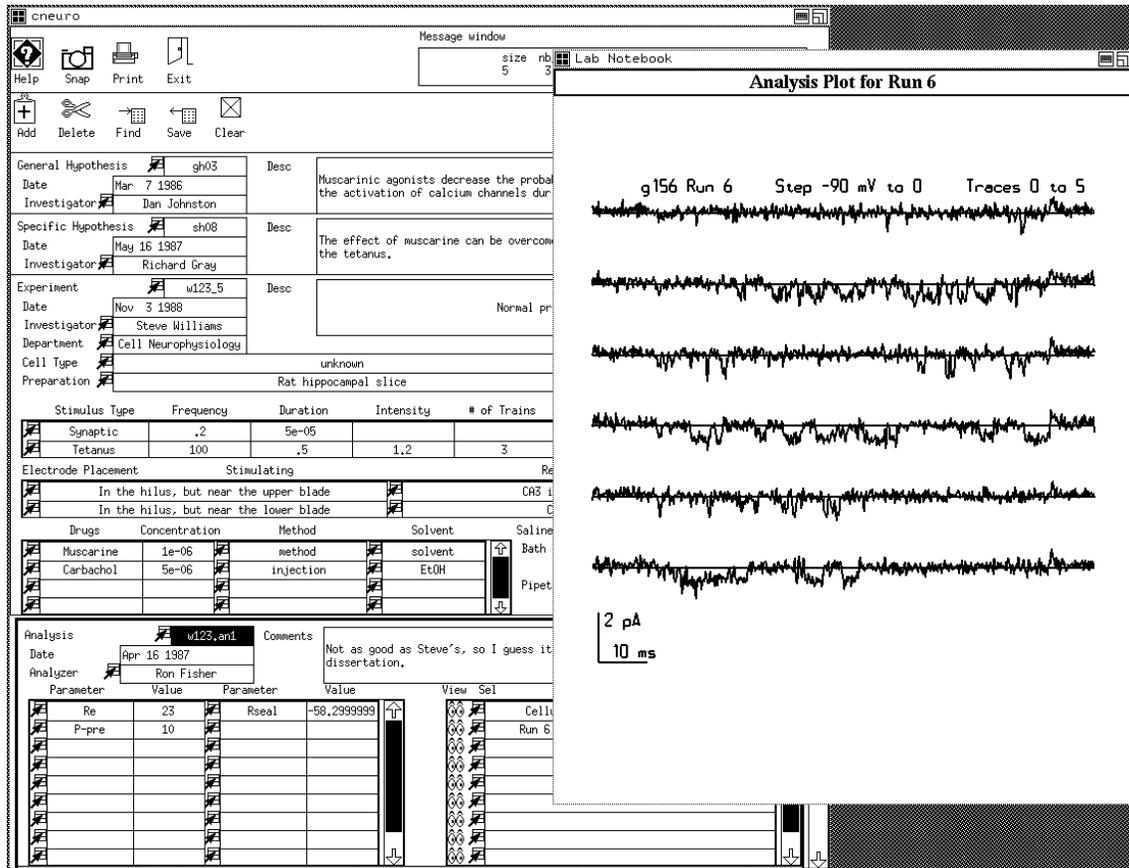


Figure 7. - A Specialized Application which makes use of the VNS

Our integration of these many diverse information resources through a hypertext system, to our knowledge, is unique. We feel that the hypertext paradigm is particularly appropriate for this task, because of the ability to contain and organize information in many different forms. Currently the system only supports image and text data, but future versions could contain audio, video, and animation as well.

Navigation

Using hypertext to support large work groups provides some problems that have not been addressed satisfactorily by previous hypertext systems. With a group of eight to ten people all sharing one or more notebooks and probably maintaining individual notebooks as well, the problem of finding information is apt to be acute. The typical hypertext method of navigation, namely following embedded icons, is not satisfactory. [3]

We provide a "filtering" mechanism for pages, objects, and links to support directed navigation by the user. Malone uses the concept of filtering in describing his Information Lens system, "Even though the term [filtering] has a literal connotation of leaving things out,

we use it here in a more general sense that includes selecting things out of a larger set of possibilities." [4] Such filtering in the Lens systems is analogous to the information scanning, sorting, and disseminating that is so important in biomedical groups.

In the hypertext, filtering involves searching a set of objects or pages to find those that meet a defined condition. These conditions are likely to use the attribute/value tuples associated with all objects, pages, and links. Subsequently, a user can search for occurrences of a given kind of page or object (such as "Find all pages on which there are electron micrograph images."). Filters can be formulated as database queries for content and context. The substance of the data and the structure of the hypertext network may be used as part of the description for filters. Filtering will be especially helpful in work groups where often members will be looking for information they did not organize.

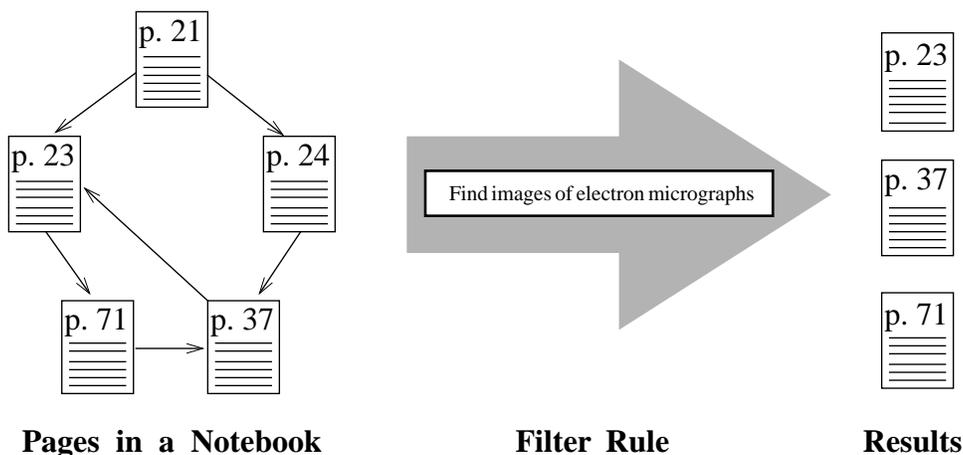


Figure 8. - An Example of a Filter Rule Acting on a Hypertext Web

Besides the typical hypertext browsing mechanism and the directed navigation of filtering, we have provided a hierarchical browser for the hypertext, that provides an outline of page titles for a notebook. The browser initially shows the notebook user's home page title. By choosing a page title with the mouse, another level in the outline appears. This level contains the titles of the destination pages of the links in the chosen page. The user also has the ability to bring up the actual page of the notebook from within the browser. The hierarchical browser is similar in function to the Browser Card in the NoteCards system, except that the detail (or depth) of the browser is interactively controlled. The browser also allows the user to search for page titles and to mark pages for easy access later. The hierarchical browser allows a variation on the traditional hypertext browsing ability in that the browser allows following links without the cluttering of the screen with the whole pages. This is useful for traversing large networks where the user has some familiarity with page content. Figure 9 shows the hierarchical browser with a notebook page.

Conclusion

The Virtual Notebook System is intended to support work groups in the biomedical setting. The need to integrate the system into a heterogeneous environment has dictated a number of the design features. The system has been organized in a manner to facilitate the gradual progression of the VNS towards a general purpose system. Much thought has gone into the initial architecture to allow future work in the areas of document preparation, version-

ing, and the addition of dynamic objects and links. We believe that through separating the data and the interface and by using a distributed relational database to store the data, we will be in a strong position to adapt the system design to these and other emerging needs.

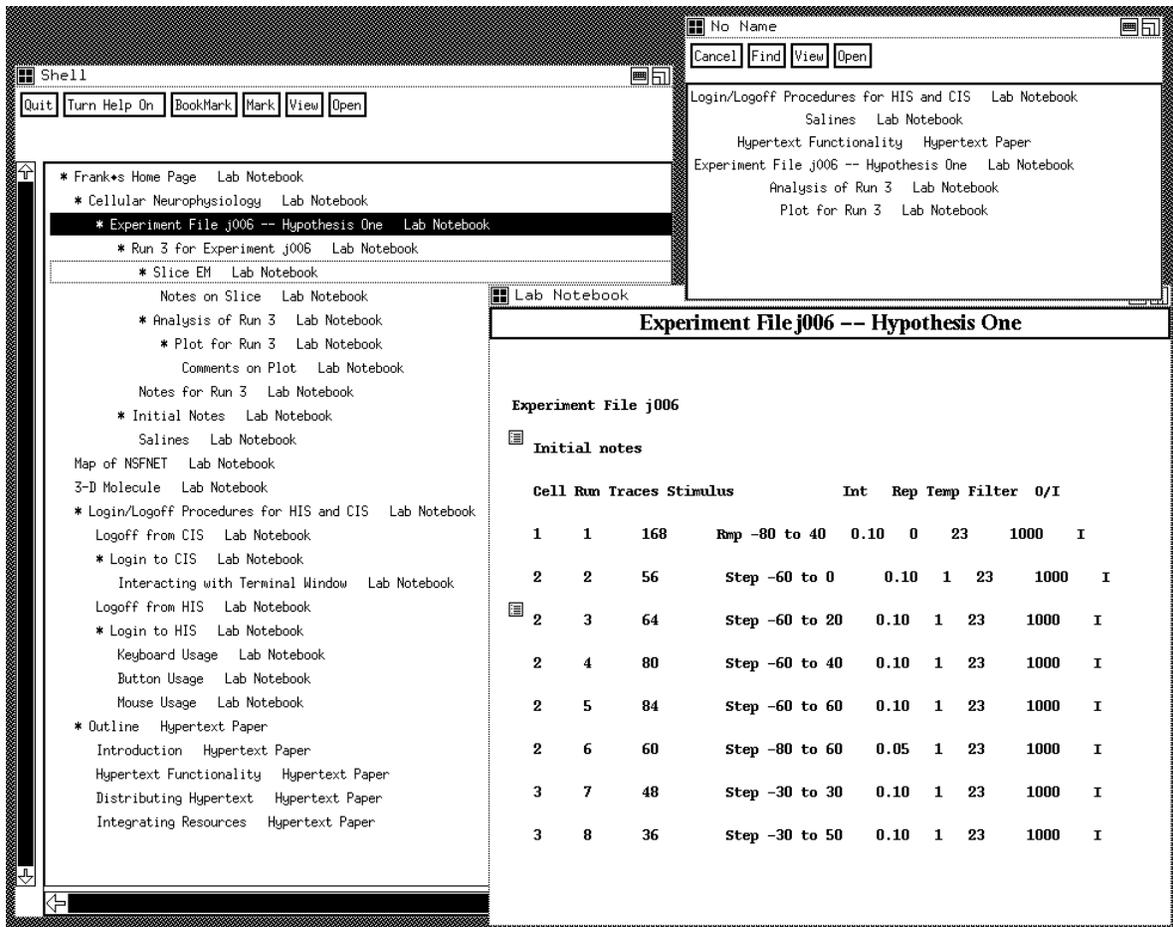


Figure 9. - The Hypertext Browser and a Notebook Page

The Virtual Notebook System is particularly well suited for the research environment, providing an easy to use method for sorting and sharing the large amount of varied data and ideas typically found in the work group setting. Colleagues may easily annotate shared information with notes, questions, or more data, promoting greater domain expertise. By integrating a set of resources that can directly insert information into the Notebook, the system relieves the user of the necessity of learning the many different communication protocols to access the many information resources he may wish to use. This reduces the reluctance that many users have to utilizing new resources and improves the efficiency of retrieving desired information. Our use of hypertext to integrate information resources in a consistent environment makes the VNS difficult from the other hypertext systems being reported.

Acknowledgment

The work reported here was supported in part by Grant LM040905-01 from the National Library of Medicine. We wish to acknowledge the work of Stan Barber, Andrew Burg-

er, Norman Furlong, Kevin Long, and Eric Taylor. Several of the applications discussed in this paper are the result of their work.

References

1. Gorry, G.A., et al., A Virtual Notebook for Biomedical Work Groups, *Bulletin of the Medical Library Association*, Vol. 76, No. 3 (July 1988), pp 256-267.
2. Conklin, E.J., Hypertext: An Introduction and Survey. *Computer*. 20, 9 (Sept. 1987), pp. 17-41.
3. Halasz, F.G., Reflections of NoteCards: Seven Issues for the Next Generation of Hypermedia Systems. *Communications of the ACM*. 31, 7 (July 1988), pp. 836-852.
4. Malone, T.W., et. al. Intelligent Information-Sharing Systems. *Communications of the ACM*. 30, 5 (May 1987), pp. 390-402.

DEC is a trademark of Digital Electronics Corporation.

GRATEFUL MED is a trademark of The National Library of Medicine.

HP is a trademark of Hewlett Packard Company

IBM is a trademark of International Business Machines, Inc.

KMS is a trademark of Knowledge Systems, Inc.

Macintosh is a trademark of Apple Computer, Inc.

MS DOS is a trademark of Microsoft, Inc.

NoteCards is a trademark of Xerox Corporation.

SAS is a trademark of SAS Institute, Inc.

Sun Workstation, Sun 3/60 and Sun 3/280 are trademarks of Sun Microsystems, Inc.

Sybase is a trademark of Sybase, Inc.

UNIX is a trademark of AT&T Bell Laboratories.

X Window System is a trademark of Massachusetts Institute of Technology.