


for every block and census tract in the U.S., as well as U.S. foreign trade data by commodity from every city in the U.S. to every country in the world. It also contains the digitized map outline boundary data for city blocks for the entire U.S. (census TIGER files).

The data are stored on 45 Pioneer six-disk CD-ROM jukebox changers connected to four Sun Microsystems SPARC workstations. Users can access the data in four ways: (1) Census LOOKUP, a World-Wide Web server providing unrestricted access to a 30% subset of the data (1990 Census data). As of the end of 1994, LOOKUP had been accessed by 6,000 different users. (2) A menu system for MS-DOS computers having NFS Internet capability; this method has had more than 17,000 recorded accesses. (3) Direct NFS (Network File Services) mounts. (4) Anonymous FTP (file transfer protocol).

The URL <http://cedr.lbl.gov/cdrom/doc/cdrom.html> describes all four access methods.

#### Acknowledgments.

This work was supported in part by the Director, Office of Epidemiology and Health Surveillance; Office of Health; Office of Environment, Safety and Health; U.S. Department of Energy under Contract No. DE-AC03-76-SF00098. 

---

*Deane Merrill* is a computer scientist at Lawrence Berkeley Laboratory (LBL) and coprincipal investigator of the Populations at Risk to Environmental Pollution (PAREP) project. *Nathan Parker* is a software designer at LBL and a student in Electrical Engineering at University of California, San Diego. *Fredric Gey* is assistant director of the UC Data Archive and Technical Assistance, which manages social science and health statistics databases for the University of California, Berkeley. *Chris Stuber* is an employee in the Telecommunications Division of the U.S. Bureau of the Census.

---

© ACM 0002-0782/95/0400

## *Envision: A User-Centered Database of Computer Science Literature*

**Lenwood S. Heath, Deborah Hix, Lucy T. Nowell, William C. Wake,**

**Guillermo A. Averbach, Eric Labow, Scott A. Guyer, Dennis J. Brueni,**

**Robert K. France, Kaushal Dalal, and Edward A. Fox**

**P**roject Envision is an early NSF-funded digital library effort to develop a multimedia collection of computer science literature with full-text searching and full-content retrieval capabilities. Envision was launched in 1991 in accordance with the ACM Publications Board's plans for encouraging research studies to develop an electronic archive for computer science.

The Envision system consists of specially-developed components: a query server, an object-oriented database management system, a presentation server, and the Envision client. The query server is a vector-space search system (i.e., MARIAN [2]). The object-oriented DBMS deals with the many kinds of entities and media types involved. Since bibliographic and full-text entries are encoded in SGML, the presentation server converts retrieved documents to HTML for final rendering. Hence, Envision's document delivery components can behave as a WWW server, though considerably higher functionality results from coupling with the rest of this highly interactive system. In particular, the Envision client provides a much more powerful user interface than is commonly found with retrieval systems.

Envision features flexible information visualization, displaying search results as a matrix of icons. As shown in the figure, each document in a search results set is represented by an icon in the Graphic View, while the Item Summary shows a textual listing of bibliographic information. The Graphic View gives users control over how search results are displayed, by manipulating values assigned to the matrix axes. Users can select the interpretation of icon position along the x- and y-axes, as well as icon color, shape, size, and label. Document attributes that can be shown in the display include relevance, publication year, type, size, author name, and index terms.

This interface resulted from extensive studies of users and their tasks, careful design and implementation to support user visualization and interaction needs, and iterative refinement in concert with usability studies [3, 4]. Extensive empirical studies show strong user interest and satisfaction.

The ACM data in the Envision database include bibliographic records from *ACM Guide to Computing Literature*, review articles from *Computing Reviews*, full-texts supplied by the typesetter of the

*Transactions*, and full-texts keyed in (less figures and mathematics) of *Communications*. These were all converted to SGML (Standard Generalized Markup Language) records to be processed. Other ACM-supplied data include CD-ROMs from the Design Automation Library, several versions of the Hypertext Compendium, and videotapes of various ACM conferences. When no electronic representations were available, pages were scanned and stored as TIFF files.

Among the key principles of the project [1, 3] was to develop an archive of objects, with SGML used as the representation scheme for both bibliographic and full-text entries. Accordingly, a tool was developed to assist with SGML Document Type Definition (DTD) preparation. This facilitated devising a DTD for bibliographic records and one for ACM *Transactions* articles. These were passed on to ACM and incorporated into the DTDs being used for the ACM digital library (see the articles appearing in this issue about ACM publishing plans, policies and guidelines). The DTDs were also used to define the target representation for converting ACM-supplied data into SGML. Since the volume and variety of materials requiring conversion were so great, we developed an automatic document analysis and translation system to produce correct SGML versions of incoming documents.

Envision has paved the way for an NSF-funded project on "Interactive Learning" (see <http://ei.cs.vt.edu/Elproj.html>), to investigate how a digital library can improve computer science education. As part of this study, software and data will be provided to ACM, along with a computer to run the Envision system.

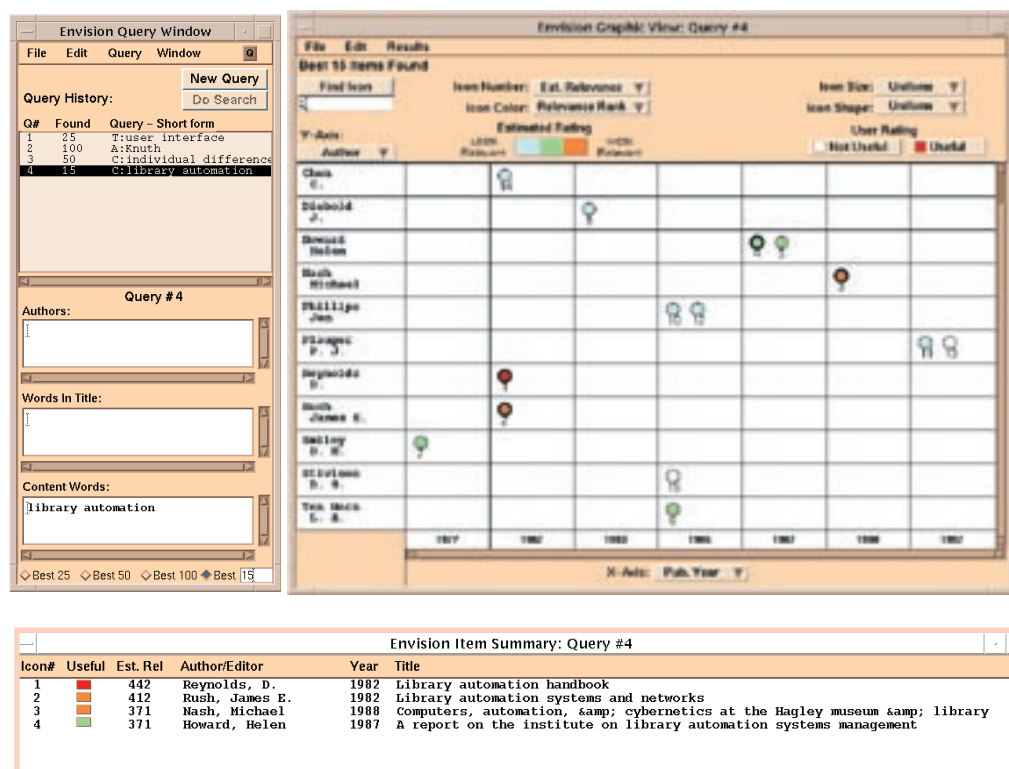
### Support

NSF funded "A User-Centered Database from the Computer Science Literature" at Virginia Tech for

1991–1995, and ACM allowed free use of its publications. Principal investigators are E. Fox, L. Heath, and D. Hix. □

### References

1. Fox, E.A. How to make intelligent digital libraries. In *Proceedings of the 8th International Symposium on Methodologies for Intelligent Systems*. (Charlotte, NC, Oct. 1994). *Lecture Notes in Artificial Intelligence 869*, Springer-Verlag, Berlin, 27–38.
2. Fox, E.A., France, R.K., Sahle, E., Daoud, A., and Cline, B.E.



**Figure 1.** Envision user interface designed by Lucy T. Nowell and Deborah Hix

Development of a modern OPAC: From REVOLC to MARIAN. In *Proceedings of the 16th Annual International ACM SIGIR Conference on R&D in Information Retrieval*. (ACM, New York, 1993) pp. 248–259.

3. Fox, E.A., Hix, D., Nowell, L., Brueni, D., Wake, W., Heath, L., and Rao, D. Users, user interfaces, and objects: Envision, a digital library. *J. Amer. Soc. Info. Sci.* 44, 8 (Sept. 1993), 480–491.
4. Nowell, L., and Hix, D. Visualizing search results: User interface development for the Project Envision database of computer science literature. In *Proceedings of the Human-Computer Interaction Conference International 1993*. Elsevier, New York, pp. 56–61.

The authors are all associated with the Department of Computer Science, Virginia Tech, Blacksburg, Virginia 24061-0106.