

PowerBookmarks: A System for Personalizable Web Information Organization, Sharing, and Management

Wen-Syan Li^{*†} Quoc Vu[†] Edward Chang[†] Divyakant Agrawal[†] Kyoji Hirata[†]
Sougata Mukherjea[†] Yi-Leh Wu[†] Corey Bufi[†] Chen-Chuan Kevin Chang[†] Yoshinori Hara[†]
Reiko Ito[†] Yutaka Kimura[†] Kazuyuki Shimazu[†] Yukiyoshi Saito[†]

1. Motivation

Users of the Web are overloaded with information. This medium is "polluted" with redundant, erroneous and low quality information. A WWW survey of 11,700 users conducted from April 10 to May 10, 1996[1] indicates that 30.31% of the users report "finding known info" is their problem and 27.80% of the users report organizing collected information as their problem. An empirical study[2] on users' revisit patterns to WWW pages found that 58% of an individual's pages are revisits. With these study results, we believe the Web users would like to build and organize a larger collection of bookmarks for future references than they can reasonably maintain now.

2. System Architecture

PowerBookmarks is being developed under the auspices of a larger WebDB project to address the above issues. WebDB hypermedia database system[3] is built on top of NEC PERCIO Object-Oriented Database Management System[4]. Unlike most search engines, which focus on information retrieval based on keywords, WebDB aims at supporting database-like comprehensive query functionalities as well as navigation on document structures, contents, and linkage information to support more advanced functionalities. WebDB provides storage, object-oriented document modeling, SQL3 query language, and HTML/VRML document generation for *PowerBookmarks*. Two external components, JTOPIC full text search engine[5] and WordNet[6], are used for full text search and on line lexical dictionary reference for the tasks of indexing and query expansion. WQL[7] is used for querying WebDB. A visual query interface is

supported to assist users in specifying queries while actual WQL queries are generated automatically by the WQL query generator. Hence, the complexity of the underlying schema and the query language remains transparent to the user.

PowerBookmarks supports information sharing with access control. *PowerBookmarks* supports automated bookmark classification based on document *contents* through an external classifier. *PowerBookmarks* also provides many useful personalized services, such as automated dead link and inactive link removal and new or modified document subscription services. A proxy server is used to monitor users' behavior to provide automated bookmarking for frequently accessed URLs. *PowerBookmarks* utilizes the query, modeling, and navigation capabilities provided by WebDB while augmenting them with information sharing, access control, and personalization functionalities. We illustrate their relationship in Figure 1.

3. PowerBookmarks Functionalities

In this section, we summarize some major functionalities of *PowerBookmarks* as follows:

Querying the Internet and Bookmarks: *PowerBookmarks* supports a simplified query interface for users to query the Internet. After the results are returned by Internet search engines, a filter program extracts useful information and presents it to the user. The user can select a set of URLs to browse in a slide-show fashion or save them as bookmarks in *PowerBookmarks*. When the user requests to collect URLs into *PowerBookmarks*, *PowerBookmarks* performs a sequence of tasks as follows: (1) downloading the URLs; (2) extracting keywords, summary, links, and related URL metadata, such as last modified date, etc.; (3) indexing the URLs in TOPIC and WebDB. After the documents are indexed in *PowerBookmarks* (through WebDB), the user can issue more complex queries to retrieve bookmark URLs. Figure 2 shows the query interface used in *PowerBookmarks*. In this window, the user issues a query retrieving the bookmarks related to call for papers for conferences related to *XML*. Queries with more complex criteria, such as links, full text search, relaxed keyword search[8], are also supported. The query functionalities in *PowerBookmarks* is carried out by the

¹Corresponding author. Email: wen@ccrl.sj.nec.com

²C&C Research Laboratories, NEC USA, Inc., 110 Rio Robles, M/S SJ100, San Jose, CA 95134, USA

³5th Development Department, 2nd Computers Software Division, NEC Corporation, 1-10, Nisshin-Cho, Fuchu-Shi, Tokyo 183-8501, Japan

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGMOD '99 Philadelphia PA

Copyright ACM 1999 1-58113-084-8/99/05...\$5.00

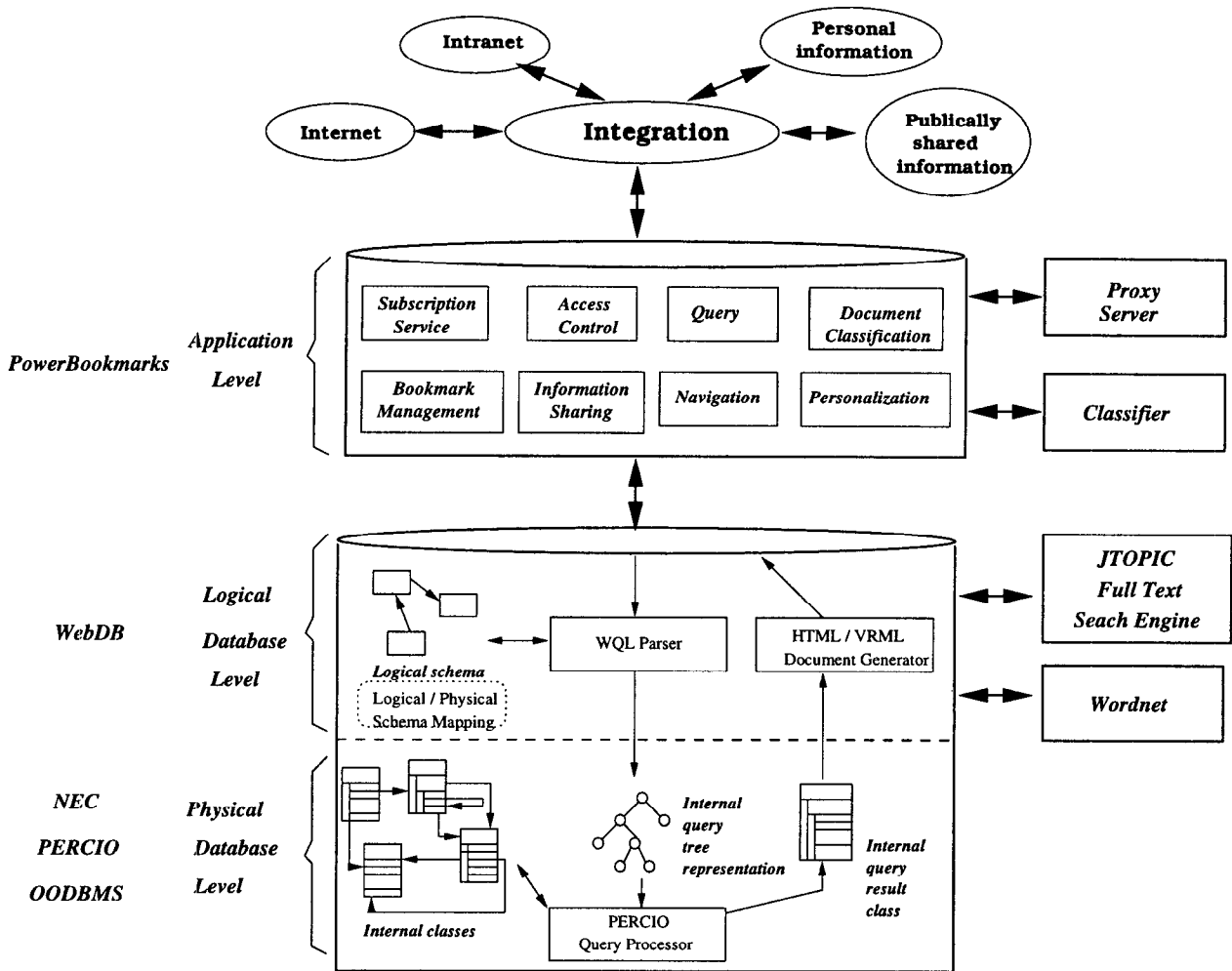


Figure 1: *PowerBookmarks* Architecture and Functionality Overview

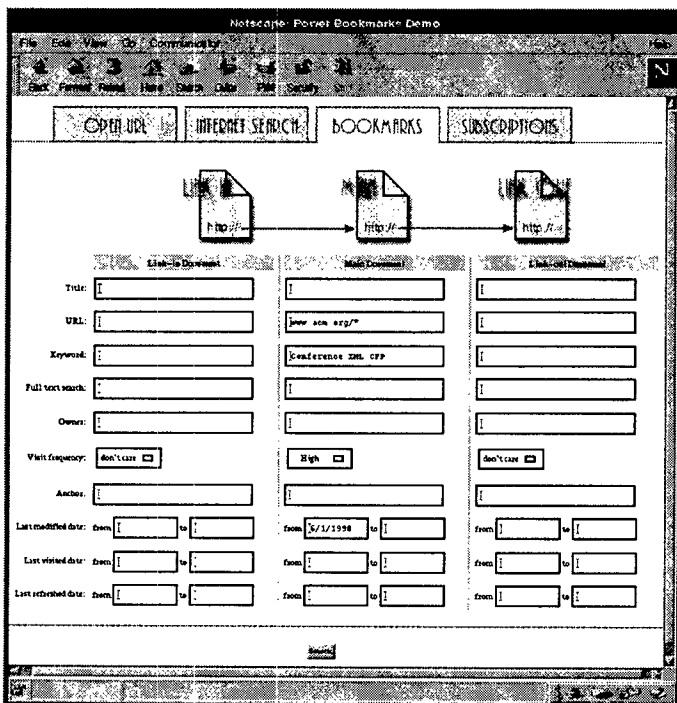


Figure 2: Query Interface

underlying WebDB.

Classifying Bookmarks for Navigation: *PowerBookmarks* supports automated document classification by contents using an external classifier similar to Pharos[9]. Figure 3 shows a classification tree for bookmark URLs in *PowerBookmarks*. Note that this window also displays the number of bookmarks in each folder (i.e. category) and its access frequency indicated with a temperature icon. The user may navigate through the classification tree to locate URLs of his/her interests. Figure 4 shows a window dump for display of documents and folders at the node *Computer/Software*. At this node, there are the user's own bookmarks (shared and non-shared) and shared bookmarks owned by other users. The bookmarks with a red cross are dead links detected by the system. The system can also automatically, if the user prefers, move the dead links and inactive bookmarks to the *Deleted folder* at the bottom of the window. The user can click on any folder to go to other nodes in the bookmark structure.

Subscription Service: *PowerBookmarks* supports subscription services for new or modified documents on the Internet and intranet. The user can set the subscription

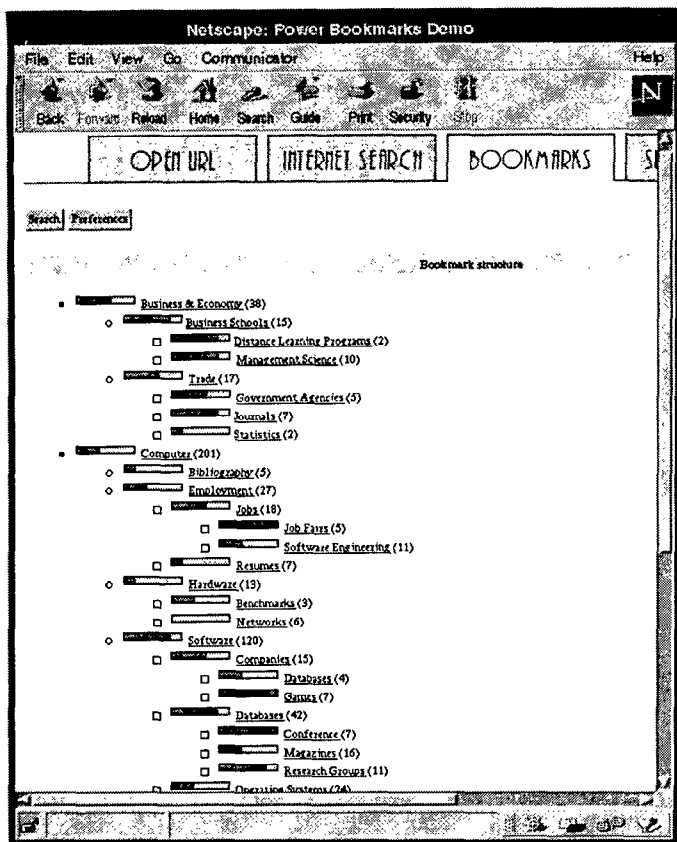


Figure 3: Bookmark Structure

criteria, such as temporal, domains, keyword similarity, or document similarity, etc.

4. Conclusion

This demonstration highlights many features of an advanced and personalizable Web information management system. Complete description of this system is available in [10] and at <http://www.ccr1.nec1ab.com/webdb/>.

References

- [1] Georgia Tech Research Corporation. GVV's 7th WWW User Survey. Information available at http://www.gvu.gatech.edu/user_surveys/survey-1997-04/, 1997.
- [2] Linda Tauscher and Saul Greenberg. Revisitation Patterns in World Wide Web Navigation. In *Proceedings of the 1997 ACM CHI Conference*, Atlanta, GA, March 1997.
- [3] Wen-Syan Li, Yi-Leh Wu, Junho Shim, Kyoji Hirata, Sougata Mukherjea, Divyakant Agrawal, Yoshinori Hara, Reiko Ito, Yutaka Kimura, Kazuyuki Shimazu, and Yuki-yoshi Saito. WebDB Hypermedia Database System. *IEICE Transactions on Information Systems*, E82-D(1), January 1999.
- [4] NEC Corporation. *PERCIO/SQL User's Guide*. NEC Corporation, 1996.

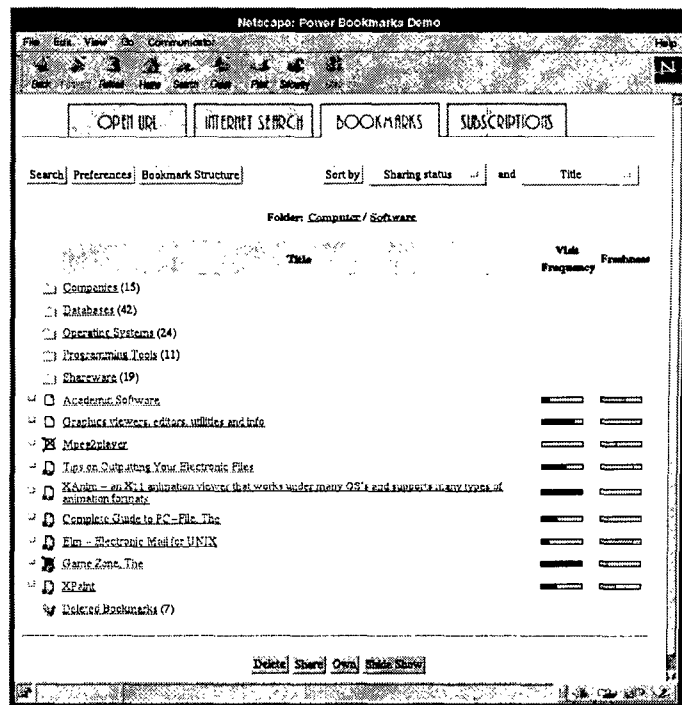


Figure 4: Navigation through Classification Hierarchy (Root → Computer → Software)

- [5] NEC Corporation. *JTOPIC Developer's Kit*. NEC Corporation, December 1997.
- [6] G. A. Miller. WordNet: A Lexical Databases for English. *Communications of the ACM*, pages 39–41, November 1995.
- [7] Wen-Syan Li, Junho Shim, K. Selçuk Candan, and Yoshinori Hara. WebDB: A Web Query System and its Modeling, Language, and Implementation. In *Proceedings of the 1998 IEEE Advances in Digital Libraries Conference*, Santa Barbara, California, USA, April 1998. IEEE.
- [8] Wen-Syan Li and Junho Shim. Facilitating Complex Web Queries through Visual Interfaces and Query Relaxation. In *Proceedings of the 7th World-Wide Web Conference*, Brisbane, Queensland, Australia, April 1998.
- [9] R. Dolin, D. Agrawal, A. El Abbadi, and J. Pearlman. Using Automated Classification for Summarizing and Selecting Heterogeneous Information Sources. *D-Lib*, Information available at <http://www.dlib.org/dlib/january98/dolin/01dolin.html>, January 1998.
- [10] Wen-Syan Li, Quoc Vu, Divyakant Agrawal, Yoshinori Hara, and Hajime Takano. PowerBookmarks: A System for Personalizable Web Information Organization, Sharing, and Management. In *Proceedings of the 8th World-Wide Web Conference*, Toronto, Canada, May 1999.