

Proposing A Private Virtual Library

5 R - 9

Mayumi Ueda, Shinichi Ueshima, Shuzo Yajima
Graduate School of Informatics, Kansai University

1 Introduction

As the network society grows rapidly and globally, we have more chances to look for information over networks. In most cases, we do not know much about information sources; where they are, what data are stored in them, how they are stored, we often have troubles in finding what we are really looking for. In addition, necessary information we want to use is often distributed over distinct information sources, and we have to gather these information with appropriate techniques, e.g., sequential access to information sources, and agent technologies, and so on.

We propose *PVL*, a private virtual library or a personal virtual library which would realize virtually a library of user's dream. PVL is a virtual environment where a user can collect information over the Internet according to a user's preferences, read them, organize them flexibly. We employ and extend the data warehousing technology to realize a personal information repository. We present the overview of PVL, and discuss some relevant issues.

This research aims at developing intelligent information media in the advanced networking environment, where PVL is suggested[1].

2 PVL: Private Virtual Library

Features

PVL is a digital library over networks for personal use, which retrieves, stores, and reuses information according to a user's preferences. A user can get information over network in the same way as reading a book from his private bookshelf. We summarize its features as follows.

- **Reading Virtual Books and Real Books**
In PVL, a user can read *virtual books*, which are retrieved over networks. He/she can read them as we read *real books* in our bookshelf without noticing the the existence of networks. Necessary data are gathered appropriately and dynamically from information sources, digital libraries and Web sites. PVL deals with *contents data* in various types of multimedia, and the bibliographic data as well.
In many libraries, various data are often restricted or closed for access due to the management problem and/or its safety. We believe we can avoid such problems and share those data in public by storing contents electrically.
- **In-time Access to Necessary Information**
The retrieved data can be classified into two categories. One is the data for temporary use, and the other is the data for further use or permanent use. In PVL, a user can select the

retrieved data interactively, and store them in his/her private library. The system automatically stores *meta-data*, such as location data, schema information, which enables to eliminates the inefficiency to retrieve the same data over networks repeatedly.

- **Private Library for Personal Use**
PVL contains *user profiles*, which reflects a user's preferences of favorite data, favorites information sources, types, etc. By using user profiles, the system collects those information a use might need from appropriate information sources on behalf of him/her with simple requests. Hence PVL is a private and user-oriented information system, which enables users to make customized queries based on their preferences.
- **Flexible Store and Reuse of Information**
A user can select the retrieved data interactively, and classify and store them dynamically. Besides virtual books, which are retrieved over networks, we store *real books*, which a user possesses, and *published books*, which a user do not possess, but publishers hold. These data are stored in different categories.
In PVL, a user can store raw data in virtual book, location data, books only publishers hold, books not yet published, etc. For the collected data, a user can organize them according to various dimensions to meet various purposes. These include authorwise sorting and fieldwise sorting, dynamic and multiple classification, creating views incrementally, etc.
In PVL, a user can store a single data into multiple categories without duplication. This technique is especially useful in handling invaluable archives, ancient documents and writings, etc.
- **Interactive Operations through User Interface**
A user can read multiple books simultaneously, and can add comments and/or bookmarks, compare books and make references between them. Especially, a user can specify portions of contents data in PVL, and store them for further use.
Furthermore, a user can customize his/her own user interface for comfort and friendliness. By utilizing preference data about music, the system can choose and provide the music of the day for a user.
- **Self-maintenance Capability**
PVL has the self-maintenance mechanism to retain its appropriate size for a user's interactive operations. Some data a user seldom accesses

for a specified period are to be deleted from PVL and moved into "Trash Box".

3 Architecture of PVL

We sketch the architecture of our system below. PVL consists of mediators, wrappers, graph-based flexible information base, user interfaces, and user profiles, and so on.

With mediators and wrappers, PVL has an independent architecture from those of information sources over network in data models, query systems. A user can query multiple autonomous, heterogeneous sources, and reuse the retrieved results[2].

- **Mediators**

We build a PVL by using wrappers as an interface of PVL to target sources over networks. By this, we can construct a private library which are independent from data models and query languages of source information sources. PVL is not affected by the system changes of information sources.

User-oriented processing of mediators are as follows. Mediator accepts a user's retrieval requests from use interface, customizes them by referencing user profiles, and sends them to wrappers. It, also, gives back a user those data that are obtained from information sources. It filters and deletes redundant data from the results, and handle the data according to the user profile. By this, it provides data according to a user's preferences[4].

- **Wrappers**

A wrapper transforms a request from mediator to the format of information source, and sends retrieval message to it, while it accepts the retrieval results to the format of PVL, and send back to a mediator.

- **Information Base for Persistency**

The information base has a graph-based repository for flexibility[3]. A user can organize the collected information dynamically and incrementally based on users' viewpoints. By virtue of wrappers, a user can integrate objects in heterogeneous information sources in PVL. The evolutions in our information base, such as volume, date as well as schemas, hyper references, are monitored.

In PVL, we store raw data, pointers, integrated data, and not necessarily confine to materialized views as in [2]. We also store both *query objects* and *anchors objects*, which guarantee wide use of contents data over networks.

Each user has his/her own space as subgraphs in the information base, while common space are shared among groups of users, or all users.

We realize the information base using an object-oriented database.

- **PVL Viewers and User Environments**

PVL viewer is an interface between information base and users. The system holds multiple PVL viewers in order to support each user. The information base is graph-based, and viewers extract personal data by queries and direct pointing over graph views. Each user accesses information base through viewers. PVL viewers also help user to give multiple views of the data in information base, and handle interactions among users.

- **Customizing Queries and Filtering via Profile**

User profiles are used for customizing queries a user submits and filtering the retrieved data given by wrappers. They hold history of user's queries and data used in interactive operations. The size of the history is restricted. User profiles helps to expand given queries to obtain unexpected good results, while they help to reduce unnecessary data over network.

We are making an experiment to see the effectiveness of user profiles for merging ranked data retrieved from multiple sources, where we evaluate the results by setting scores[4].

4 Concluding Remarks

In this paper, we proposed a framework of private virtual library. We believe this system is applicable and extensible in various fields, since PVL has (1) an independent architecture from those of information sources over network in data models, query systems, etc., (2) information base itself is flexible, (3) the features of user-oriented system that it utilizes feedback from each user. The application of PVL to the Internet is straightforward. *PVM* or a private or personal virtual museum is also studied as having the similar architecture mentioned in the paper.

References

- [1] Yajima, S., "Kandai Triangular Educational Network", Kansai Univ., Int'l Symposium, Informatics-The Future of Information Society, Kansai Univ., Sep.1998.
- [2] Widom, J., "Research Problems in Data Warehousing", Proc. of 4th Int'l Conference Information and Knowledge Management (CIKM), 1995.
- [3] Ueshima, S., Morishita, J., Ohtsuki, K., and Sugiyama, T., "Incremental Data Organization of Semi-structured Data Using Hierarchical Graph Model", Trans. IPSJ, Vol.39, No.4, Apr.1998.
- [4] Sho, M., "Customizing Query with User Profile: Utilization of Feedback from Users", Graduation Thesis of the Faculty of Informatics, Kansai Univ., Feb.1998.