Blog-based research notebook: personal informatics workbench for high-throughput experimentation

Shin-ichi Todoroki *, Tomoya Konishi and Satoru Inoue

*Advanced Materials Laboratory, National Institute for Materials Science,
Namiki 1-1, Tsukuba, Ibaraki 305-0044, JAPAN*

Abstract

In this age of information technology, many researchers are still conservative in keeping a log of their activities in paper-based notebooks. This style of log-keeping brings about the situation that our experimental data and their descriptions are recorded separately into hard disks and papers, respectively. Such a data separation is likely to be a serious rate-limiting factor in high-throughput experimentation from the view point of getting feedback on each researcher's work from what he has done. We propose to utilize a blog (Weblog) as an electronic research notebook and discuss technical requirements for maintaining it, on the basis of the blogging experience for four years by one of the authors. We need a user-installed blog server with authentication function for personalization and network infrastructure enabling us to "blog anytime, anywhere". Although some knowledge-sharing systems have similar electronic notebooks as their front-end, the present blog system is different from these because it stores personal information which is not meant to be shared with others. This blog-based notebook cooperates with these e-notebooks by promoting hyperlinks among their contents, and acts as a personal informatics workbench providing connections to all the resources needed.

Key words: blog, informatics, electronic notebook, high-throughput experimentation
PACS: 07.05.Kf, 07.05.Wri

* Corresponding author.
  Email address: `TODOROKI.Shin-ichi at nims.go.jp` Tel:
+81-29-860-4348 Fax: +81-29-854-9060 (Shin-ichi Todoroki).
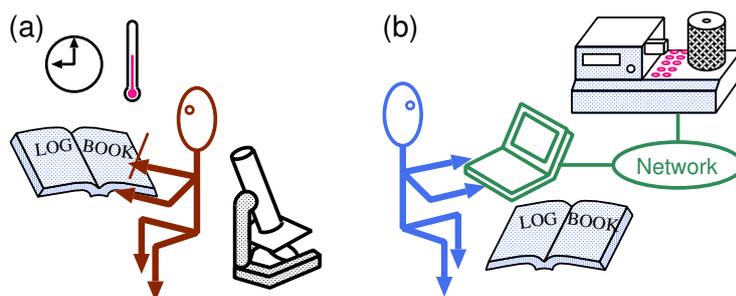 *URL:* `http://www.geocities.jp/tokyo_1406/` (Shin-ichi Todoroki).

Fig. 1. An illustration comparing two styles of logging researcher's activity, old (a) and new(b).

1 Introduction

In this age of information technology, many researchers are still conservative in keeping a log of their activities in paper-based notebooks. This style of log-keeping involves one problem, which we have overlooked for a long time since the widespread proliferation of personal computers (PC); our experimental data are separately recorded into our hard disks apart from our logbook which stores the rest of information such as experimental strategies, data evaluation, supplemental notes etc. (see Fig. 1). This separated data storage is likely to be a serious rate-limiting factor in high-throughput experimentation because of its bad traceability of the context of past experiments; we have to consult both our PC and logbook at the same time and to reorganize the information by the aid of some keywords of dates, filenames and other knowledge in our patchy memories. Such a task prevents us to get a quick feedback from past experiments, especially for the case of treating large amount of data.

The core of this problem is that the information stored in paper-based media needs manual operations for data retrieval and referencing to other information, in contrast to the ones stored in electronic media. The solution to this problem is very simple. All information should be stored electronically in one place and discard paper-based logbooks. It is, however, easier said than done.

Here we propose to utilize a "blog" as an electronic research notebook since a blog system provides user-friendly interface compatible with web browsers, easy-to-use authoring tools and full-text retrieval. A blog-based research notebook is different from original blogs in one important way; its readers are restricted to its author (and his/her colleagues, if necessary). We discuss technical requirements for maintaining it, on the basis of blogging experiences for four years by one of the authors. In this paper, after summarizing the blog technology briefly (section 2), we discuss the requirements to use it as our research notebook (section 3) and give tips to enhance the traceability between our experimental data and blog-based research notebook (section 4). Finally, we compare the present notebook with existing electronic notebooks[1,2] (Section 5). Such a discussion is useful not only for researchers but also

for the system vendors willing to carry on their business in informatics tools.

2 Popularization of blog

The word, blog, comes from "web log", a publicly accessible personal journal for an individual. This is similar to a personal diary, but is shared with other people over the web. Blogs appeared in late '90 and became popular since some free services for blogging appeared, such as http://www.blogger.com/ and http://blog.goo.ne.jp/.

Basic functions of blogs are the followings; (1) displaying the contents in order of time, (2) displaying specific contents, responding to user's requests of dates, topics and keywords, (3) registering newly edited contents including graphic arts, multimedia and links to other resources such as past blog contents, web pages, local data files and remote databases, and (4) all the functions are called via web browsers without any preliminary knowledge.

3 Requirements as our research notebooks

Since a blog is basically a communication tool, its contents are to be open to the public, but it is not appropriate for our research notebook. Thus, user authentication function is to be added. Secondly, the contents should be browsed and updated anytime and anywhere during our research activities. Lastly, all the resources to be included in the notebook should be accessed via network wherever you are browsing it. In this section, technical requirements to realize the above features under the existing blog systems are discussed.

3.1 User authentication

Since this feature is generally not available in the existing free blogging services, we have to set up at least one computer acting as a web server including a blogging service. The software packages needed are a HTTP (Hypertext Transfer Protocol) server and CGI (Common Gateway Interface) tools for blog service. The authentication function is provided by the HTTP server. There are variety of user-installed blog software packages (for example, see [3]) and we can find some which meet our specific requirements, such as cost, operating system, skills etc. The authors constructed a Linux-based system including Apache HTTP server[4] and tDiary[5].
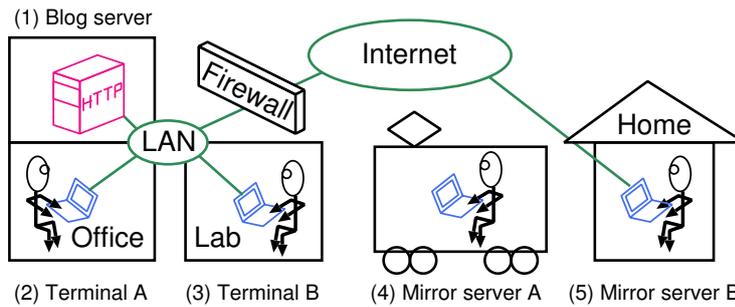
Fig. 2. An illustration showing accessibility of a blog server from its clients.

3.2 Accessibility to the blog server

In order to realize "blogging anytime, anywhere", networked terminals need to be available in the office, laboratories, and at home. If the server has to be accessed over a firewall blocking HTTP requests, it is necessary to place a mirror server outside the firewall (see Fig. 2(5)). The next best solution is to construct such a mirror server in a notebook-size computer and carry it along when venturing outside a firewall(see Fig. 2(4)). Wireless access via PDA (Personal Digital Assistant) or cellular phone can be preferred by people who do not care for carrying them all the time and their restricted usability due to their compactness.

3.3 Accessibility of the contents

Since the contents of the research notebook are sent over the HTTP protocol, all the data to be included, such as text, photographs, figures and links to other resources, have to be stored in HTTP servers including the present blog server and other servers speaking the HTTP protocol. The latter involves database servers for a closed user group, managed by research projects in order to share data among their members (for example, see [1,2]). All the contents mentioned above are embedded in the main text of the research notebook by HTML (Hypertext Markup Language). Figure 3 shows an example of blog output.

4 Enhancement of traceability between a blog and data files

Once this blog-based research notebook is set up and kept for some time, we can retrace the past experimental data from this blog notebook as long as we have embedded them in it (see see dashed arrows in Fig. 4 pointing to (2) from (1)). However, it is more frequent for us to refer in the opposite direction, that is, we refer to the page of the notebook corresponding to the date on which the data we are seeing were collected (see dashed arrows in Fig. 4 pointing to (3) from (2)). It is therefore
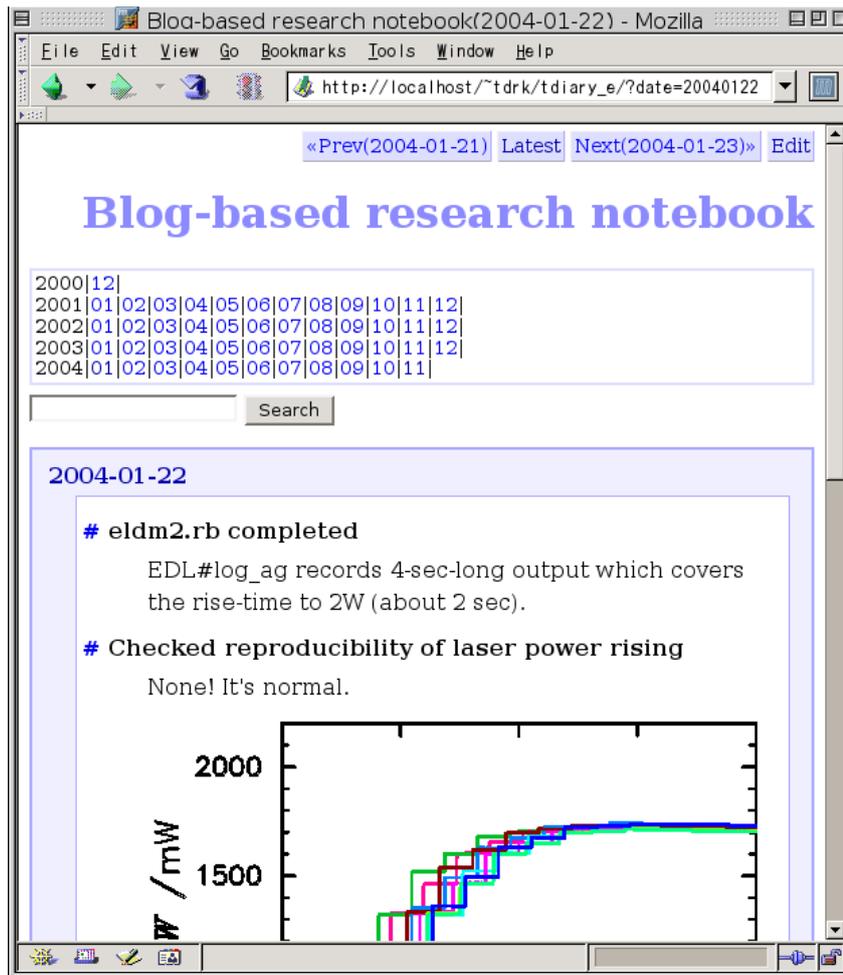
Fig. 3. A screenshot of a blog-based research notebook managed by one of the authors.

necessary to have an automatic means of producing such references instead of typing the date in the blog every time it is needed. This function has to be provided by the software for data-visualization and also the database servers mentioned above.

We developed a simple program to visualize our data and to open the corresponding page of the logbook as shown in Fig. 5. The user can choose one datafile in the lower half of the window. This software then shows the date of data storage in the upper half of the window. The selected data is visualized by clicking "Visualize by" button located near the center of the window. The corresponding page of the blog notebook is shown by clicking "Display the Diary" button in the top of the window, just like in Fig. 3.

By the aid of such a utility, we can access all the information through our PC without any stress for managing separated data storage as mentioned in Section 1 and the right side of Fig. 1.
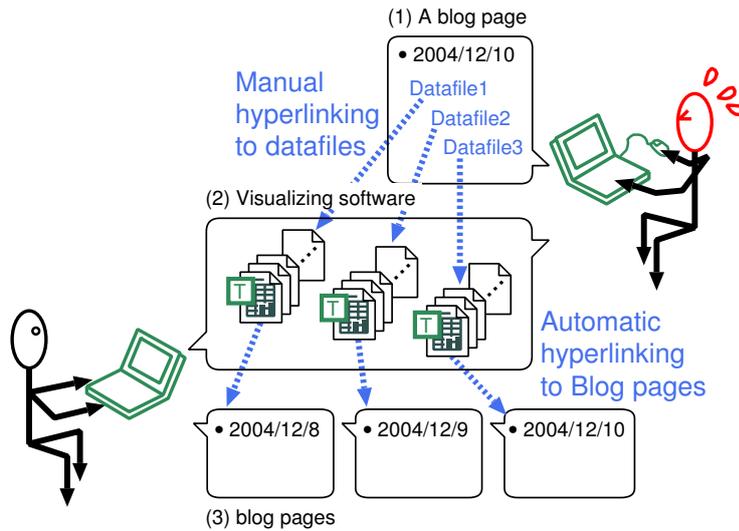
Fig. 4. An illustration showing hyperlinks between blog pages and data files. Rounded rectangles are pages of the blog-based research notebook or visualizing software. Dashed arrows are hyperlinks between these pages.
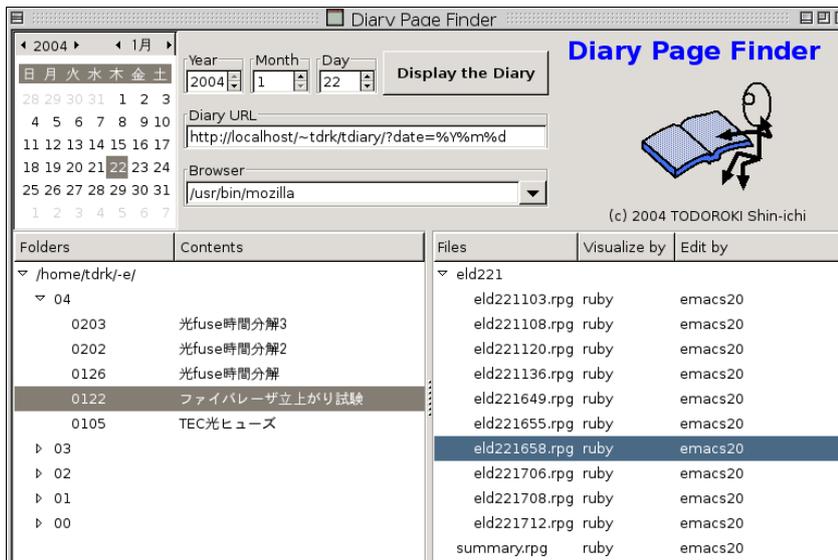

Fig. 5. A screenshot of a utility program for visualizing datafiles and showing corresponding pages of the blog-based research notebook. It is written in an object-oriented scripting language, Ruby[7], with Ruby-GNOME2 library[8].

5   Comparison with former electronic notebooks

Until now, several "electronic notebooks" have been proposed as a front-end of knowledge-sharing systems[1,2]. There is no big difference between these notebooks and the present blog-based notebook from a viewpoint of functions of presenting the stored data. They are different, however, in practical use. The contents of the former are to be shared among the users of the knowledge-sharing systems.

6

The users upload to the server only what other users are interested in. On the other hand, the present blog stores every information the blog owner will refer to in the future, that is, what we used to store in a paper-based research notebook.

A practical advantage in the blog-based notebook is that each individual can start blogging in their own fashion even if existing electronic notebook service is not available. The people who are already using an electronic notebook provided by one of the knowledge-sharing systems can find benefit in starting their own blog-based notebook even if they worry about another separated data storage. As long as each of their data servers allows the contents to be linked to each other, the user does not feel the border between different systems. These notebooks should be properly used depending on the purpose; for data sharing or for private use. In addition, a sophisticated technique to make a crosslink between two blog servers is very popular in the blog community, called "trackback"[6]. It is possible to make a link to specific page in any page of a remote blog server accepting trackback through CGI. This function is essential for existing knowledge-sharing systems in order to cooperate with their users' blog systems.

6 Conclusion

Blog-based research notebook enables us to manage all the personal information electronically, which was formerly recorded in our paper-based notebooks. This service requires user authentication function in a user-installed server and infrastructure for "blogging anytime, anywhere". Once these are satisfied, the blog acts as a personal informatics workbench; a gateway to all the information needed which is traceable and retrievable. Although some existing knowledge-sharing systems also have electronic notebook service as their front-end, the present and existing notebooks should be properly used depending on the purpose; managing personal information or promoting knowledge sharing.

References

[1] S. J. Ludtke, L. Nason, H. Tu, L. Peng, W. Chiu, Object oriented database and electronic notebook for transmission electron microscopy, Microscopy and Microanalysis 9 (2003) 556–565.

[2] S. Meguro, T. Ohnishi, M. Lippmaa, H. Koinuma, Elements of informatics for combinatorial solid-state materials science, Meas. Sci. Technol. 16 (1) (2005) 309–316.

[3] O. Winkler, Blog software breakdown, (http://www.asymptomatic.net/blogbreakdown.htm) (Apr. 2004).

[4] The apache software foundation, (http://www.apache.org/).

[5] tDiary, (http://sourceforge.net/projects/tdiary/).

[6] Mena and Ben Trott, A beginner's guide to trackback, (http://www.movabletype.org/trackback/beginners/, Japanese translation: http://kotonoha.main.jp/weblog/000138_trackbackguide.html) (Mar. 2003).

[7] Ruby Home Page, (http://www.ruby-lang.org).

[8] Ruby-GNOME 2, (http://sourceforge.net/projects/ruby-gnome2/).