A Proposal of a Personal Information Management System based on searching and clustering

Ryo UMETSU

School of Knowledge Science,
Japan Advanced Institute of Science and Technology
September 2001

Keywords: PIM, information management, searching

Now, It's a new century, which is an information age. The many informations are there around us these days. But we tend to fail to manage these informations well. On the other hand, as Alvin Toffler said in his book "Power Shift", knowledges come to play a important role in the new era. So it's needed to manage informations around us and produce knowledges. Now I propose a new personal information managing system, which helps to manage a lot kinds of informations efficiently, and to reuse and reproduce knowledges.

It's true that many PIM tools are already found anywhere. But most of them aren't efficient managing system, because some of them have a poor searching function, or others are hard to install, etc.

For example, the memoma is a free software on MS windows. It's a convenient application for taking memo. However, re-using written informations

in it is awkward. Searching functionality is not smart, and it's difficult to look over the search results.

Fusen-shi 2000 is also a windows application. It provides a yellow(default) Post-It paper like text window. A user will write memo down on it. But like a physical Post-It, it's easy to be buried by other application windows. And users can't search through Post-It papers.

Another example, Q-Pocket, made by Mr.Masui at the Sony CSL, has a good concept. Its data are based on a text file type, a file for a new memo is automatically named by the unique time date number. Q-Pocket is, however, hard to install and use, and it's a technically unrefined system.

So I propose a new personal information managing system here, named HC. HC uses text file base data, which means easy to handle from any OS or applications. It gives a new unique file name when one inputs some information. Inputting data needs no categorization to annoy users. If the system needs categorization when users input new information, users will be confused by vague categories. HC can avoid such a problem.

HC can be used from various client applications. It's enabled by a server-client system. HC also has portability. Its server is written in ANSI C. It can work on Solaris, Linux, and MS Windows.

The client program is now written in GNU Emacs Lisp, because heavy PC users are in front of editor(especially GNU Emacs) very often. Most of the time they use an editor, so when a good information occurs to them or is found, they can easily use the HC client and input it.

On reproducing knowledges, search function is available. Most of PIM applications have easy access to input data, but there's no idea on searching data. That users can search all data easily is one of the big points of this tool. This searching will help users to re-use or reproduce their informations.

Now I'll show how to use. An example usage. In the case that you're editing some files on GNU Emacs, a nice thing occurs to you. Or you search something on the WWW and get some important information. Then you'll type M-x "hc-new" to input a new memo information to the HC. A new window will open and you can input strings. After editing a new idea, the command by typing C-x C-s will end up in the inputting.

After that, when you'll need some information which you entered before, you'll use M-x "hc-search" command. You can get a search result list in the Emacs buffer. In the list, memos that include search keywords are listed up chronologically. Like most MUA(Mew, Wanderlust, mh-e), typing "n" and "p" will shows the contains of the memo on the list line. You can browse

memos including search keywords very easily.

In the list window, “e” means edit mode in which you can edit an existing memo. The “D” command deletes unwanted memo, “s” command invoke another searching, “q” is quitting the HC.

You can use this system like this and re-use your informations that are input before very easily.

In the future, this tool can be developed as below. First, a search result listing algorithm can be more refined. I’m thinking to use clustering algorithm like LSI or so to calculate similarities among search results. And because this is a personal system, the system will find some feature from input data. For example, the system will be able to find some bias in the input terminologies. If the system can use this bias efficiently, user searching will be realized more smoothly.

Moreover, except for critical personal data, user will be able to open memo data to the internet. If so, this system can make client-to-client connection like GNUTella, then everyone using this system will search and look over other users’ data. It’s good idea to connect to the WWW. But such case, the search results will increase heavily, clustering results or something will be needed.

On its client application, combining to a MUA is one of good options, because heavy PC users generally always open a MUA during they working. And a MUA already has functions to search, browse files, renumbering file name, etc. If this HC system can connect(or parasite) to such a MUA, this enables users to use seamlessly with mail system. Mail and memo system will be unified.