

Snippets of Awareness: Syndicating Copy Histories

M. Cameron Jones

Michael B. Twidale

Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

501 E. Daniel St. Champaign, IL 61820 USA

{mjones2, twidale}@uiuc.edu

ABSTRACT

We explore the potential of syndicating copy actions between group members, particularly as a way to provide lightweight activity awareness information. The advantages of the approach include its ease of integration with existing activities and tools, its flexibility and its lightweight nature for both giver and receiver of information. Its disadvantages include the small amount of information captured limiting its interpretability and usefulness for full awareness, and regular concerns about privacy. RSS syndication allows for a rapid exploration of alternative design ideas and easy integration of activity awareness with commonly used information awareness tools.

Categories and Subject Descriptors

H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces - *Computer-supported cooperative work.*

General Terms

Design, Human Factors, Documentation.

Keywords

Copy-Paste, Shared Clipboard, Activity Awareness, Syndication

1. INTRODUCTION

In earlier work on collaborative help-giving [9] we had noted how people often had many applications open at the same time. They could be viewed as constructing complex, flexible and, where needed, ad hoc workflows. This involved copying and pasting items from one application to another. Sometimes this was a matter of assembling elements from several sources into a final document, and sometimes it was more complex, such as copying results from a number of databases into a spreadsheet, doing some calculations, creating a graph, copying that into a drawing package to improve its appearance, copying the result into a report and emailing it round for comments and updates.

These workflows were not explicitly represented nor directly supported. They had to be managed by the users themselves. They are undertaken by many kinds of people including those who claim to have little competence or confidence with computers. They are a widespread phenomenon, easily recognizable and with several reports in the literature. Indeed they are so widespread as to be in danger of being considered mundane and beneath the attention of CSCW research, involving commercial off the shelf software with little explicit collaborative support built in. And yet they are successful in getting the collaborative work done. They also enable people to cope flexibly with challenges like new work requirements, systems failures and gaps in understanding the full power and functionality of all the systems employed [9]. Consequently, we find copy-paste intriguing in its power in

enabling end user creativity in appropriating and combining applications to address complex tasks. We have been considering how the power and ubiquity of the approach can be exploited and enhanced. In this paper we describe an initial exploration of the implications of collecting, storing, representing, syndicating and hence sharing copy-paste activity across all the applications used in a work task. In particular we consider the potential of this information as a lightweight awareness mechanism with additional possibilities for supporting ad hoc workflow creation and coordination.

2. SYNDICATED CLIPBOARD DESIGN

The Syndicated Clipboard tool has three main components: a clipboard logging application, an HTTP web server, and an RSS reader and clipboard loader.

2.1 Clipboard Logger

A small, continuously running background application monitors the user's clipboard and records all the content cut or copied to it (the privacy features are described in section 4.7). Each text item from the clipboard is saved into a separate file and made accessible to the web server for distribution in the RSS feed. An unintended consequence of this configuration is that the clipboard files are then searchable by the user through a desktop search application like Google Desktop or Microsoft Windows Search.

2.2 Web Server

In our prototype, the syndicated clipboard also contains a simple HTTP server which publishes the most recent clipboard events in an RSS feed. The default feed contains the ten most recent copies; there is also a feed for viewing the entire clipboard history. The current built-in web server does not utilize any user authentication mechanism, security, or encryption. However, an operating system-level firewall could be used to actively restrict access to particular subnets, domains, and IP addresses.

2.3 Client

The client-side consists of an RSS client and a small clipboard data loader, registered in the Windows Registry. By using a popular open standard like RSS, the clipboard data can be viewed from a number of available tools, many of which may already be embedded in the users' current activities. As illustrated in the figures, we have tested various RSS readers. In addition RSS aggregator services like SuprGlu and NewsGator can construct meta-feeds of multiple users' clipboards, perhaps useful for small group collaboration. Each entry in the syndication feed contains a link field with a URL to the single clipboard item. When the user clicks on the link for an entry in their RSS reader, the clipboard data loader fetches the clipboard data from the origin and copies it into the user's local clipboard. Figure 2 is example of the clipboard content loaded by the data loader. All figures are based on data created while writing this paper, with some details altered.

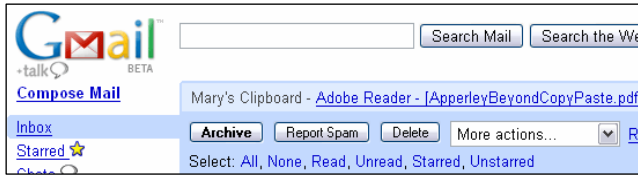


Figure 1 Mary's Clipboard feed embedded in John's Google Mail account

3. RELATED WORK

The Microsoft Office Clipboard supports individual use, showing a list of recent copy events along with an icon indicating the application that was the source of the copy action.

Loosely-coupled collaborative tools, including shared clipboards, afford spontaneous and lightweight collaboration, employing low-costs for initiation and overhead to facilitate communication and information sharing [8]. Patel and Kalter believe that coordination in loosely-coupled systems is difficult and best handled outside of the system. Likewise, Miller and Myers [7] reported on coordination problems in synchronized clipboards, finding that users had difficulty tracking the contents of the clipboard when shared among multiple users for periods of time.

The act of copying as a trigger point for giving a system awareness of a user's focus of interest and thus enabling the provision of context-sensitive assistance or functionality has received some attention [1].

Fenstermacher & Ginsburg [4] considered how the recording of large amounts of client-side user activity with different applications can provide a far richer understanding of website use than is possible with just server-side web log analysis. However this approach suffers from the problem of how to analyze so much data, and how to obtain it from people who may be unwilling to sell their privacy for minimal immediate reward. Although copy-paste data may prove a terse but easy to collect data source for subsequent analysis, in this paper we are considering its immediate use to those who share the data, not those who may subsequently analyze it.

4. DISCUSSION

4.1 Lightweight Development

The relatively low cost of development has allowed us to begin exploring the design space of syndicated clipboards by trying out different RSS readers, data structures and functionalities. The ease of incorporating existing functionalities and services allows a far more free-form ad hoc exploration of a possibility space than we have encountered in conventional systems development which requires substantial efforts to create the first working prototype, and consequently more careful planning and initial design.

Using RSS as the protocol for syndication allowed for the rapid trial of numerous clients, many of which were already installed on our computers and embedded in our daily workflows. Leveraging existing systems in this fashion led to the discovery of a

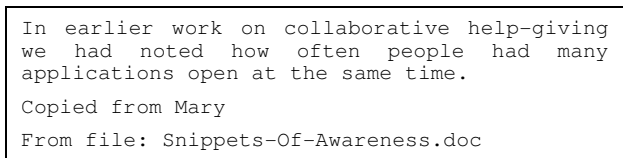


Figure 2 Example of syndicated clipboard content

lightweight awareness character to having the feeds available in the applications already in use.

4.2 Lightweight awareness

The importance of facilitating awareness, particularly in remote collaboration is a significant theme in CSCW research [2]. Lightweight awareness mechanisms that have been developed by end users in physical co-located settings have been noted in many workplace studies, and can include paper use, muttering, or even looking at the output of a shared printer [5].

We believe that syndicated copy-paste data offers at least an interesting potential as an awareness mechanism. Like all others it involves various tradeoffs [5]. We did not design it to be an awareness mechanism. We were more interested in its potential for helping individuals and their colleagues gain a sense of the nature of their cross-application activity in order to facilitate helping and information sharing. Awareness, if it should turn out to be effective, is an added bonus. The main advantage of syndicated copy-paste for awareness is its ease of collection. The users do not have to do any additional work, other than to consent to their data being recorded (see later), and it is relatively easy to collect and syndicate. Its terseness means that potentially a long sequence of activities can be skimmed easily. But is it too little or too much information to be useful? Our preliminary exploration of using the tool ourselves indicates at least some promise.

The ease of trying different RSS readers allows us to explore a range of visualization options. In Figure 1 the clips feature in GmMail is used, allowing only one snippet to be viewed at a time with scrolling required to view them all. Figure 2 shows use of the Firefox's Live BookMark, giving a richer view of several snippets in a list, whereas the Google Desktop Sidebar, also in figure 2 provides richer expansions of the information at the cost of more screen real estate. Figure 4 shows the use of Internet Explorer 7, giving the full details in a single view. Figure 5 shows FeedReader, a stand-alone, email-like reader, integrating with a user's other news feeds. The user can see timestamps on the feeds, as well as applications and filename information. Each format has its own tradeoffs, but it can be left to the end user to integrate with their own preferred RSS viewer. None of them has much meaning without wider knowledge of the work context. However, if Mike were collaboratively writing a document with Mary, seeing her snippets would have at least some meaning to him.

It is certainly possible at the cost of more programming effort to collect more information and try to communicate more about the work context. For example, in our current implementation we only record copy events. Would it help to also record paste events and to reconcile this data to show how data was copied from one application to another? We will consider that in future work.

4.3 Coordination

Our original interest in copy-paste was in how to support and enhance its use in workflow creation by individuals and groups. We wondered whether a more explicit representation of how multi-application use might help in a variety of ways:

- It would help the individual keep track of what they were doing. Distractions inevitably occur and it can be difficult to remember where one was in a complex process, not least one that involves many steps that evolve depending one has arisen in the earlier steps.

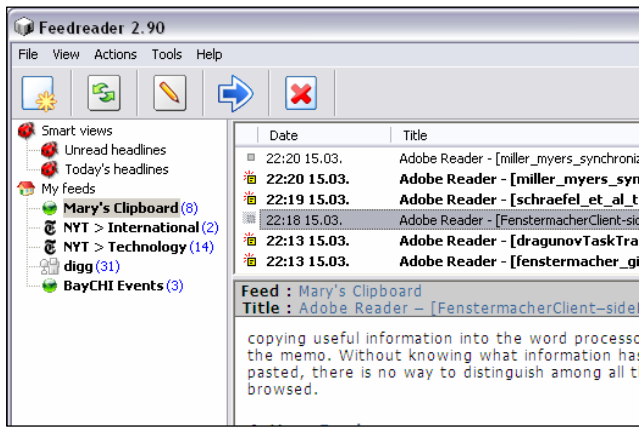


Figure 3 Feedreader, a stand-alone RSS client, showing Mary's clipboard feed displayed with other news feeds

- If the user encountered a difficulty and needed to ask a colleague for help, this would provide a lightweight and rather terse history mechanism. It would be short enough to get a sense of the inter-application context of the help request (the user may be stuck about doing something in say MS word, but understanding the larger need may mean that the best solution is to do an earlier step in Excel differently).
- If the user is doing a task similar to one done earlier, the previous sequence can serve as an explicit template for action, rather than the user having to rely solely on their memory. This may enable faster and smoother operation. Our design assumption informed by prior work is that the tasks to be supported rarely involve an exact replication of precisely the same sequence of action steps, but are similar enough to earlier tasks for a template to be a good starting point [3].
- Successful sequences may be shared with colleagues, either as standalone templates or as an externalized crib to be used as a basis for explaining the sequence used.

4.4 Granularity of recording and visualizing

Fenstermacher & Ginsburg were chiefly interested in logging all user activity in order to gain a better understanding of web usage in context. We too could aim to record far more user interaction. This imposes greater costs of development, of storage (the authors' usage of the text-only, clipboard logger generated an average of 50 kilobytes of text per day) and of devising visualizations so that the data is actually of use, particularly to end users rather than analysts (the main focus of Fenstermacher & Ginsburg's work). There is also a greater cost in privacy – we can imagine users being less willing to let every menu choice (or even every window context switch) be recorded than to let only copy actions to be recorded. Assuming that there are contexts where these wider issues are successfully addressed, effective activity visualization will still require different levels of granularity of detail. Although currently only speculation, the promise of interpretability of a copy-only representation means that it may be that a much larger dataset may provide different scales of detail of view and that a copy-only visualization is of use in this richer information context. It does tell us a lot about certain critical inter and intra application coordination activities.

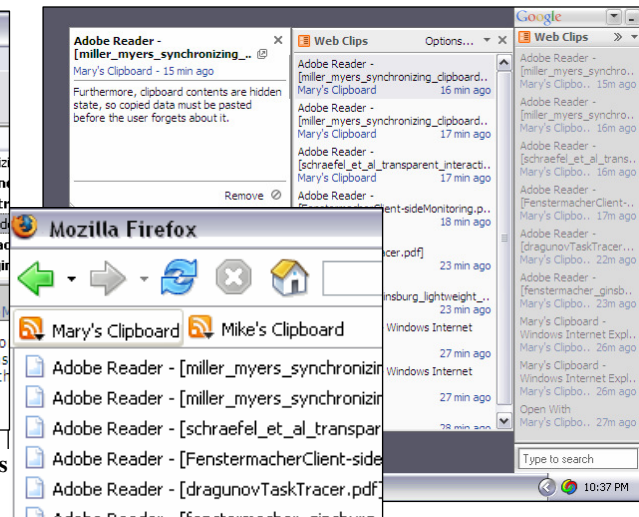


Figure 4 Firefox's Live Bookmarks' and Google Desktop's Web Clips' views of Clipboard feeds

4.5 Collaborative information retrieval

In our current use of the tool while writing this paper, we can see its power in supporting collaborative information retrieval [10] where it is desirable to gain an awareness of a variety of different aspects of a colleague's information seeking activities so as to coordinate and complement the work done. In CIR one needs to share direct results (items found that are judged relevant), but also process information: which databases have been searched, which keywords tried, which queries have yielded so much information that additional work is needed, and which have been closed off. A shared clipboard does seem to have great potential as a low effort means of sharing both process and product, assuming the participants already understand the context of the work task. As well as awareness of what the other person is doing, it is possible to directly copy results or query terms for one's own search use. It remains to be seen how well this scales to supporting awareness and coordination between more than two users.

4.6 Privacy concerns

All work on awareness has to consider issues of privacy, if only to acknowledge the tradeoff between gaining some advantages of being aware of others' activity versus some disadvantages of others being aware of your activity. The dimensions of weighing privacy concerns are complex but include issues of reciprocity (can I see you as you see me), awareness (do I remember that it is

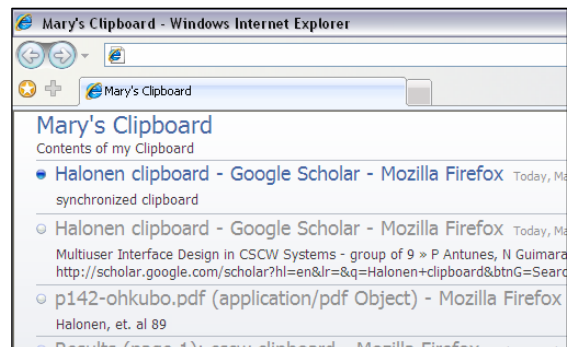


Figure 5 A sequence of copies associated with an information search

on), control (can I turn it off), usefulness (what do I get from 'selling' my privacy), clarity (do I know what I'm giving up) and committedness (can I change my mind about permissions, maybe even retrospectively). Although only a prototype that we are currently just testing ourselves, we are well aware of the privacy concerns and have built in some initial controls, well aware that others will be needed as we move to larger scale testing.

For example, a user can enable and disable logging by right clicking on the clipboard's notification icon in the Windows system tray. The clipboard logger contains a set of regular expression filters used to filter sensitive data (e.g. credit card numbers and social security numbers) from being logged and published. In light of the fact that only content actively copied by the user is recorded and published, this reduces the likelihood of sensitive information being accidentally published. Beyond filtering the content being copied, regular expressions could be used to filter specific applications, folder paths, and URLs to allow users to prevent the system from logging data copied from sensitive sources like bank websites or financial applications. A user can remove an item from syndication, by deleting the corresponding file from the log folder.

4.7 Experiences in use

We have been using the tool for just a few days, and so can only talk about the early stages of technology adoption. Installation and use is quick and easy. We are aware that our activities on our computers are now much more public (even if only shared with a single trusted colleague). Switching on and off recording is easy, but it is also easy to forget, leading to the accidental sharing of data such as snippets pasted into emails to third parties that might be somewhat undesirable. However as these are fragmentary, these faux pas are not quite as embarrassing as if the entire email was on view to another. The ability to yank back potentially embarrassing snippets is very useful. A certain time delay before enabling syndication may additionally be useful, as it is before actually sending certain controversial emails written in haste.

Against those negative points, it is rather enjoyable to share aspects of the co-authoring process without the overhead of a fully shared editor. It is amusing to consider what, if anything, a colleague will make of certain copy actions that seem meaningful only within the micro-context of the last few seconds. Equally one may occasionally copy something just to 'show' to the other person, without it being of direct use to you. This pushing of information is easy, but not necessarily noticeable. The recipient may not be aware of its intentional significance amidst all the other copies not done for show. We can imagine providing a way of emphasizing particular copies in future versions of the tool.

A sequence of copy actions can give a colleague a sense of which tasks or sub-parts are currently being attended to, and some sense of attention switching and multitasking. The sequence of copy actions as least in our preliminary use is very bursty – periods of many copy actions in a short time and then long periods of no visible activity or only occasional copies. This information offers the potential for assessing attention levels and interruptability. Naturally it also requires additional contextual knowledge to disambiguate between intense activity on one project that might be amenable to an interruption about that project, and times when interruptions are more suitable during periods of light activity. Again the advantage of this method is its relatively low impact in

terms of collection and privacy violation, but at the cost of perhaps insufficient information for making appropriate decisions.

5. CONCLUSION AND FUTURE WORK

In developing a rapid lightweight syndication of copy-paste data to explore ideas of supporting and extending the widespread use of inter-application ad hoc workflows by individuals, we realized that this approach offers great potential as a lightweight awareness mechanism. It has many advantages as a minimally invasive collaborative feature, but at the risk of providing insufficient awareness information to be widely useful. We continue to explore the possibility space to understand the aspects of designing particular applications of the idea that address classic trade-offs of privacy, granularity, ease of use and effectiveness.

Currently the tool only supports copying text. We would like to support native data types ideally anything from any application. It would be helpful to centralize the distribution scheme so that the feeds are available even when an individual is not (say, when they go offline with a laptop). We also want to consider collecting more data, such as paste events and switches between windows. Naturally we want to observe usage of the tool over longer periods of time, first on ourselves and then by more users. This will allow us to test various modes of security and permissions for acceptability and effectiveness.

6. REFERENCES

- [1] Dix, A., Beale, R. & Wood, A. (2000). Architectures to make Simple Visualizations using Simple Systems. Proceedings of Advanced Visual Interfaces. 51-60.
- [2] Dourish, P. & Bellotti, V. (1992). Awareness and coordination in shared workspaces. Proceedings of CSCW'92. 107-114.
- [3] Dragunov, A.N., Dieterich, T.G., Johnsrude, K., McLaughlin, M., Li, L., & Herlocker, J.L. (2005). TaskTracer: A Desktop Environment to Support Multi-tasking Knowledge Workers. Proceedings of IUI. 75-82.
- [4] Fenstermacher, K. D. & Ginsburg, M. (2002). A Lightweight Framework for Cross-Application User Monitoring. IEEE Computer 35(3) 51-58.
- [5] Grasso, A., Meunier, J. L. (2002). Who can claim complete abstinence from peeking at print jobs? Proceedings of the 2002 ACM conference on CSCW. 296 - 305.
- [6] Gutwin, C. & Greenberg, S. (1998). Design for Individuals, Design for Groups: Tradeoffs between Power and Workspace Awareness. Proceedings of CSCW 1998. 207-216
- [7] Miller, R. C. & Myers, B. A. (1999). Synchronizing Clipboards of Multiple Computers. Proceedings of User Interface Systems and Technologies. 65-66.
- [8] Patel, D. & Kalter, S. D. (1993). Low overhead loosely coupled communication channels in collaboration. Proceedings of ECSCW'93. 203-217.
- [9] Twidale, M.B. (2005). Over the shoulder learning: supporting brief informal learning. Computer Supported Cooperative Work 14(6) 505-547.
- [10] Twidale, M.B., Nichols, D.M., Paice, C.D. (1997). Browsing is a collaborative process. Information Processing & Management, 33(6), 761-83.