
Management of Personal Information Scraps

Michael Bernstein

MIT Computer Science and
Artificial Intelligence Laboratory
32 Vassar Street
Cambridge, MA 02139 USA
msbernst@mit.edu

David Karger

MIT Computer Science and
Artificial Intelligence Laboratory
32 Vassar Street
Cambridge, MA 02139 USA
karger@mit.edu

Max Van Kleek

MIT Computer Science and
Artificial Intelligence Laboratory
32 Vassar Street
Cambridge, MA 02139 USA
emax@csail.mit.edu

mc schraefel

University of Southampton
Electronics and Computer
Science
mc+chi@ecs.soton.ac.uk

Abstract

We introduce research on *information scraps* – short, self-contained personal notes that fall outside of traditional filing schemes. We report on a preliminary study of information scraps' nature and outline plans for the next phase of our user study. Based on ongoing study results, we describe our designs and prototypes for information scrap capture and access tools.

Keywords

Personal information management, information scrap

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): Input devices and strategies.

Introduction

Information workers today have an enormous assortment of tools and techniques - both digital and physical - to help them organize, record, and retrieve their data. In the digital realm, simple file and folder hierarchies are now accompanied by organizational strategies such as tagging, full-text keyword search, and time-based approaches. Interestingly, ethnographic research (e.g., [8]) has consistently found one kind of data to resist organization. It exhibits itself as small files strewn on the computer desktop, as Post-

Its™ littering the workplace, and as emails sent to oneself to ensure that a particular blog quote is retrievable later. We keep this information on loose paper in our pockets, save it inside other documents, and archive it with our email. This information is the subject of our research.

In this work-in-progress paper we introduce a line of Personal Information Management (PIM) research toward the design of a new tool for managing *personal information scraps*. We define information scraps as *short, self-contained notes intended for their author's use*. Information scraps typically span a few words to a few partial sentences in length, containing only enough clues to evoke the complete thought in the author's mind. They serve a variety of purposes, including functioning as reminders, memory aids, or holding places for incomplete thoughts.

Despite their widespread use, personal information scraps are not well handled by conventional tools and thus a variety of ad-hoc keeping strategies have arisen to control them. The typical information scrap requires more effort to file correctly than the user is willing to concede to the endeavor; as a result, a lightweight, catch-all backup strategy is used to ensure that the note is archived, indexed, and/or classified. While there are a number of tools dedicated to scrap management in the digital world, from Sticky Notes to BackPackIt.com, it is not clear that an optimal approach for scrap support has been developed.¹ Our research therefore steps back to fundamentally examine

information scraps: what they are, what kind of user information they encapsulate, how they are used now, and how they *might* be used if we were less constrained by current application/data paradigms. We take as a fundamental assumption that information scraps are *not* user filing failures, but instead are failures of the *interface* to support the characteristics and needs of a specific class of data. Solving the information scrap problem will help knowledge workers manage the variegated and often fragmented information they deal with in their everyday work

Related Work

Previous ethnographic work has identified how people use short, personal notes in their daily activities. Malone *et al.* [8] discovered that the “pilfers” in their study habitually kept notes on overflowing whiteboards and strategically placed scraps of paper on their desks to serve both reminder and quick-access functions. Blandford *et al.* [5] studied the use of short personal notes for task work and found that informal tools like paper and unstructured text files were sometimes preferred over traditional PIM applications because they supported more freeform input. They found that people take advantage of the ambiguity of freeform tools deliberately, sometimes merely to save themselves the effort of having to be completely explicit – for instance, by stating “lunch with Jim Tues” rather than “Meeting: Jim Hendler, Jan 3, 12:30-13:30.” Bellotti *et al.* [3] also found informal priority lists on random scraps of paper. Since the information scraps in her study were often used to keep track of mental state at the time of an interruption, her team concluded that tools to support the creation of scraps had to be “instantly on”, constantly available, and support quick, easy and spontaneous input. Hayes *et al.* [7] examined user

¹ We have published an overview of these tools at <http://people.csail.mit.edu/msbernst/papers/pim-tools.htm>

practices surrounding the recording of short important thoughts, and similarly found that day planning and task tracking were two of the most common reasons their subjects wrote notes for themselves. A number of other studies have identified “coping strategies” for users without adequate tools to deal with information scraps. The most common such strategy is to use e-mail for keeping track of one’s information scraps, often easily identified by users who e-mail themselves reminders and notes (*e.g.*, [4]).

A growing interest in “life hacking” (www.lifehacker.com) and the “Getting Things Done” philosophy [1] has inspired a number of new tools for capturing spontaneous thoughts. These include Bederson’s NoteLens notetaking tool, designed as a *flow interface* [2]: a tool which attempts to keep users “in the flow” as they work. Similarly, commercial tools such as SlipBox and NoteBook are gaining in number and popularity. Each of these tools takes an application-oriented approach to scraps: the user must call up that application or web site both to enter and to retrieve the scrap.

We have additionally been inspired by Quicksilver (<http://quicksilver.blacktree.com>), which enables a temporary command interface to be opened using a single system-wide hotkey. The user begins by typing the noun component as a short string (*e.g.*, part of a file name, a song name, an address, or a calendar entry). The user then selects the appropriate verb to perform on that noun from the list presented; an identified phone number might enable either the address book or a phone application. The entire interaction is both fast and accurate. The successes of

this tool led us to hypothesize that a similar approach might work for information scraps.

Preliminary Study

To examine the implications of a Quicksilver-like interface to scrap entry, and to learn more about the nature of information scraps, we ran an informal weeklong “Fake Computer” diary study wherein the researchers recorded into a text file any interactions they would have had, should a system as we described above actually be available. We were interested in considering: what kind of data gets recorded in this file? What kind of contextual information would be recorded if we assumed that it might be reusable or useful later on? What kind of syntax was invented to enter, retrieve, and manipulate this data?

Preliminary Findings

A sample of entries from our text files is displayed in Table 1. We discovered a variety of recording styles: for instance, Researcher 2 and Researcher 3 assiduously annotated context data, while Researchers 1 and 4 recorded only the data itself. Consistent with Blandford *et al.* [5], notes often included deliberate ambiguities such as “do ___ stuff” or a “remind me” note without any mention of when the reminder should actually occur. Commands like “open cal” were also used. Structure ranged from very orderly notes to almost unparseable text. Verbosity also varied between clearly explicated sentences and very condensed text, even within the same log.

Reflecting on the process of recording these notes generated additional insights. For example, one

Researcher 1	do crutchfield stuff kaptchuk phone? map from home to sandberg put lifehacker.com on my personal page under news f mail from kenzie re mellon wikipedia democracy f email from matt about wiki regina's phone is ...-...-....
Researcher 2	location: stata forbes cafe when: 3pm with: Julia doing: eating lunch notes: don't eat "stuffed peppers" lunch special from forbes ever again location: robot-station (office) when: 3:15pm with: no one doing: writing email to Karger and Brainstormers off to lecture at 3:30pm remind me: take my LDA/Topic model papers to meet with JT
Researcher 3	loc: danny's office working: schedule/budget writing: "if the command line were cool" who: me (waiting for david k). sched call tues noon danny cell; alarm -10m call timbl cell open cal
Researcher 4	whittaker @ at&t email stuff bellotti @ email as habitat bookmarks stuff in 2002 saul greenberg? email jenny alvin^2 re: wii party skipping cooper scratch paper resolved

Table 1. Sample entries from the researchers' self-recorded information scrap files. Some represent information to archive, while others represent "I wish" commands given to a nonexistent system.

researcher was later unable to decipher his notes' original meanings – clearly, facilitating this kind of message unpacking will play an important role in our future work. The logs tended to evolve shorthand for meaningful tokens (e.g., "Tim Berners-Lee" became "timbl"). Also, some notes were continually updated: one of the researchers would even return later to flesh out hastily-written notes. Contextual data was also written *as interpreted* rather than as it occurred, so that the hallway outside a colleague's office become simply "<colleague>'s office". Strikingly, we found that much of the information in these logs might not have otherwise been recorded on a computer.

User Study Plans

Our preliminary study suggests that our Fake Computer encouraged capture of light context (e.g., "@ office"), light syntax (e.g., "ph:", "remind me") and ambiguous data (e.g., an entry that says only "have not yet"). We wish to explore how valid these patterns may be in a more general population.

Proposed Method

We intend to gather user data from two populations: so-called "life hackers" (those who spend significant effort reorganizing their lives to best deal with their personal information) and the rest of us. We anticipate that life hackers will provide rich data, as they have already demonstrated interest in the general problem. Conversely, the general populace will allow us to get a better sense of what attributes might be necessary for those not religiously organized. We are currently evaluating two possible study methodologies. One approach will be a variant of the diary study called the Experience Sampling Method [6]. Specifically, we plan to interrupt participants at semi-random intervals and

ask them to share their last recorded information scrap in any medium, physical or digital. After they have recalled the scrap, we will conduct semi-structured interviews to discern the context and meaning of these scraps. A second approach mirrors the preliminary study, by which we will ask participants to record *all* information scraps into a single text file for a period of time. In this study we will also encourage participants to invent syntax as necessary to invoke imagined functionality from an information scrap recording system (e.g., "CHI WIP due date" to ask the system to recall the note recording this submission's due date). We will then analyze these logs and proceed by similar semi-structured interview to gain further insight.

Current and Future Design Directions

As discussions from the preliminary study continue, we have also been exploring several different information scrap interface approaches. These approaches, described below, will continue to evolve over the course of the proposed studies.

DOINGpad

Building on the text buffer and command line approach, and inspired by the Lifestreams time-centric model [7], we have implemented an early prototype facilitating information scrap capture. The DOINGpad (rhymes with *boing pad*) specializes in quick capture of context to support refinding activity. Like NoteLens [2], it is specialized for quick input of notes and thoughts. Figure 1 is a screenshot of the current iteration of the DOINGpad. DOINGpad records the following whenever the user begins to write a note: current date and time, friendly location name (from the wireless access point; e.g., "max's office"), a webcam photo of the user and his/her surroundings, and a Uniform Resource Identifier

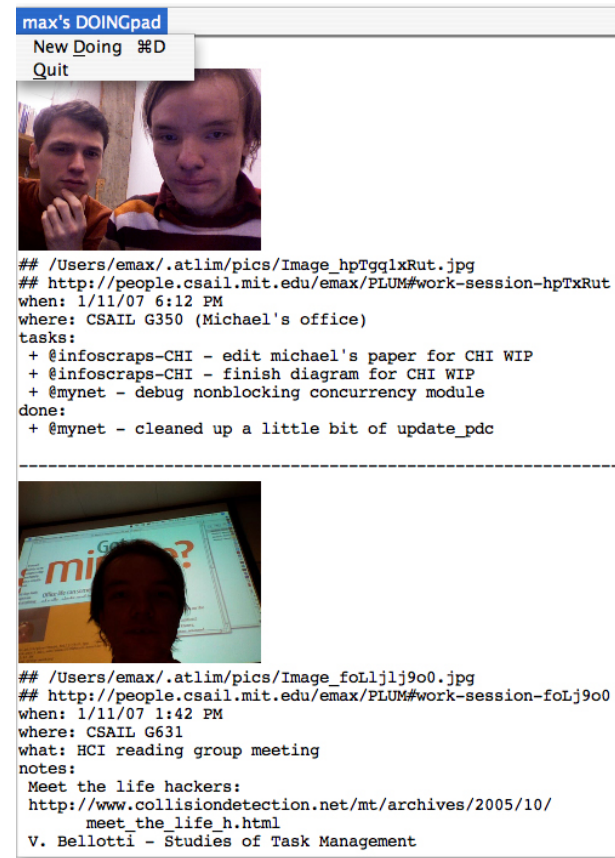


Figure 1. DOINGpad automatically captures context data such as a photo, location and time when users begin annotation.

(URI) which can be linked to other concurrent system activity such as window switches and music being played. Below these system-generated fields is a free text area for recording the note itself. DoingPad is implemented in Java and runs on Windows, Linux and Mac OS X.

Re-Finding Solutions

Where the DOINGpad examines potential input mechanisms, we also hope to examine re-finding solutions once the data has been successfully captured. We hypothesize that existing filing solutions (*e.g.*, folders, tagging, full-text search) will not satisfactorily support this activity, and hope to author an alternative design solution to support effective reminding, re-finding, and exploration. Ideally this solution may provide a means of semantic extraction to permit the construction of useful automatic views. This step, however, remains largely an open question.

Mobile Capture

As Hayes *et al.* [7] recognized, the recording of short important thoughts must not be confined to moments when the user is at his or her computer. Mobility and ubiquity are extremely important to a complete information scrap solution, and so we believe an answer may be found in mobile phones. Mobile phones have the necessary ingredients for a mobile capture solution: digital input that is always with the user. Digital (Anoto) pens present a second possible solution.

Conclusion

Information scraps – small pieces of user-generated data which fit poorly within existing filing paradigms – represent a significant percentage of our information workspace, yet to date HCI research has found no satisfactory solution for them. Scraps run counter to many of the assumptions of our current filing paradigms: for example, assumptions that all data must be given an unambiguous filing location and name immediately upon creation. (Compare performing Save As... on an internet download to slapping a hastily scribbled Post-It™ on the wall.) If research can design

support for this ad-hoc data wherein users can quickly save an information scrap *and* easily find it later, we will have more fully enabled a wide variety of information work, including note taking, memory aids, document editing and to-dos.

References

- [1] Allen, D. *Getting Things Done: The Art of Stress-Free Productivity*. Penguin Books, New York, NY, 2001.
- [2] Bederson, B. Interfaces for staying in the flow. *Ubiquity* 5, 27 (2004).
- [3] Bellotti, V., Dalal, B., Good, N., Flynn, P., Bobrow, D.G. and Ducheneaut, N. What a to-do: studies of task management towards the design of a personal task list manager. *Proc. CHI 2004*, ACM Press (2004), 735-742.
- [4] Bellotti, V., Ducheneaut, N., Howard, M. and Smith, I. Taking email to task: the design and evaluation of a task management centered email tool. *Proc. CHI 2003*, ACM Press (2003), 345-352.
- [5] Blandford, A.E. and Green, T.R.G. Group and Individual Time Management Tools: What You Get is Not What You Need. *Personal and Ubiquitous Computing* 5, 4 (2001), 213-230.
- [6] Csikszentmihalyi, M. and Larsen, R.E. Validity and Reliability of the Experience-Sampling Method. *J. Nerv. Ment. Dis.* 175, 9 (1987) 526-536.
- [7] Fertig, S., Freeman, E. and Gelernter, D. Lifestreams: an alternative to the desktop metaphor. *Proc. CHI 1996*, ACM Press (1996), 410-411.
- [8] Hayes, G., Pierce, J.S. and Abowd, G.D. Practices for capturing short important thoughts. *Ext. Abstracts CHI 2003*, ACM Press (2003), 904-905.
- [9] Malone, T.W. How do people organize their desks?: Implications for the design of office information systems. *ACM TOIS* 1, 1 (1983), 99-111.