

Personal Information Management, Personal Information Retrieval?

Michael Bernstein, Max Van Kleek,
David R. Karger
MIT CSAIL
32 Vassar Street
Cambridge, MA 02139
msbernst@mit.edu, emax@csail.mit.edu
karger@mit.edu

mc schraefel
Electronics and Computer Science
University of Southampton
Southampton, UK
mc+hcir@ecs.soton.co.uk

ABSTRACT

Traditional information retrieval has focused on the task of finding information or documents in a largely unknown space such as the Web or a library collection. In this paper we propose that the space of Personal Information Management (PIM) holds a great number of problems and untapped potential for research at the intersection of HCI and IR. In this position paper we focus on the problem of *information scraps*, or unstructured notes and thoughts, as a particularly interesting space for future research in HCI and IR.

INTRODUCTION

Information retrieval has traditionally assumed a user goal of finding information or documents within a search space whose contents may not be known *a priori* to the user, such as the Web. The user's challenge is to specify an accurate information query (e.g., "What is the capital of Uruguay?") and the system's challenge is to return the most relevant answers or documents.

Contrast this situation with Personal Information Management (PIM). Here, the user's challenge is instead to organize, find and manipulate *his or her own* information, of which the user has intimate knowledge. Though users are still performing information retrieval or search tasks, the tasks' nature may change dramatically. In PIM, users may bring much more highly contextual queries ("When is that meeting with Kerry that I set up while I was at lunch on Friday?"), and have personally authored much of the information they are attempting to retrieve. Further, the user's task does not finish with the retrieval of the document or datum; it often continues through cycles of editing and reorganization.

PIM embeds many IR tasks that users deal with on a daily basis, yet these tasks have been largely overlooked by the IR community. We believe that PIM as an area of research has much to gain from information retrieval techniques and that IR, in turn, may benefit from focusing some of its effort on PIM tasks. In this paper, we outline our research on *information scraps* [1], detailing how this PIM problem interacts closely with information retrieval, and how taking a traditional IR approach might overly decontextualize the problem.

INFORMATION SCRAPS

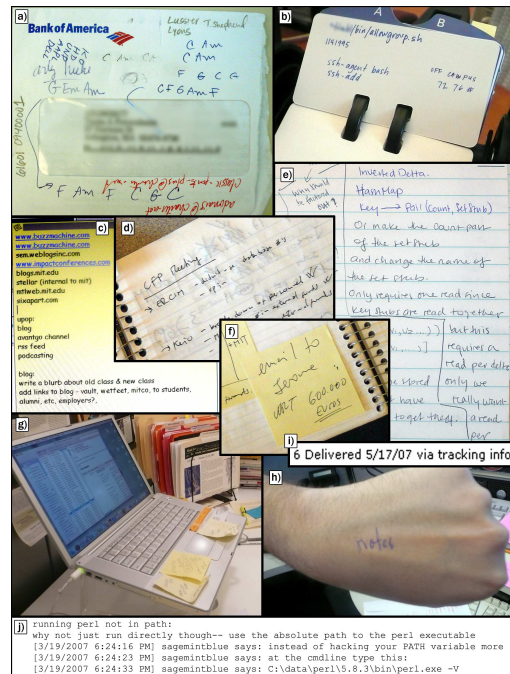


Figure 1. Information Scraps collected in our investigation of existing practice.

Despite the number of personal information management tools available today, a striking amount of our data remains out of their reach: the content is instead scribbled on Post-it notes, scrawled on corners of sheets of paper, buried inside the bodies of e-mail messages sent to ourselves, and typed haphazardly into text files (Figure 1). This scattered data contains our great ideas, sketches, notes, reminders, driving directions, and even our poetry. We refer to these pieces of personal information as *information scraps*.

We conducted a study consisting of 27 semi-structured interviews and artifact examinations of participants' physical and digital information scraps, at 5 different organizations [1]. Here, we summarize one piece of the study in support of our argument, focusing on general uses our participants found for information scraps.

Roles that Information Scraps Play

We consolidated a list of common information scrap roles from participants' responses regarding how and why they chose to store information of various types in scraps.

Temporary Storage. Information scraps' small, discardable presence enabled their common use as temporary storage. One participant kept Post-it notes on her laptop palm rest for just this purpose, recording visitors' names and contact information, later to be disposed of.

Archiving. Many information scraps were intended to reliably hold on to important personal information for long periods of time. Participants commonly used information scraps to archive notes from meetings and passwords.

Work-in-progress. Our participants shared with us many work-in-progress scraps, such as half-written emails, ideas for business plans, brainstorm, and interface designs. "Before I put anything in the computer, I like to put it on the whiteboard first," one participant explained of her newsletter layout design process.

Reminding. Many participants took advantage of information scraps' visibility and mobility by placing them in the way of their future movements to create reminders for themselves. Participants used techniques such as colored Post-its or unread or unfiled e-mails, reminding them to take action later.

Unusual Data Types. Taking advantage of information scraps' freeform nature, participants managed unique data types that might have otherwise remained unorganized. For example, one participant created an information scrap system to manage a library-style checkout for his privately owned construction tools, and one participant maintained a complex document of contact information annotated with private notes on clients.

Information Scraps, PIM and IR

Information scraps present a challenging information retrieval task. It is appropriate to consider IR with respect to scraps' lifecycle because of the sheer number of scraps our participants compiled, in tension with the need to re-find specific scraps later. However, information scraps do not easily lend themselves to traditional IR approaches. Scraps are often recorded incompletely, written tersely, or intentionally left ambiguous – making it more difficult for IR algorithms to parse the content. Conversely, the user may recall the content via a completely different set of cues than the content itself: for example, context surrounding note creation ("I wrote it while in the elevator.") or gestalt meaning ("My notes from that meeting about funding.").

Depending on the role an information scrap is playing, its information retrieval needs may further vary. For example, work-in-progress scraps may often contain very little explicit information to index – for example, consider a notebook page full of rough interface sketches. It is also questionable whether traditional IR metrics and tasks are even

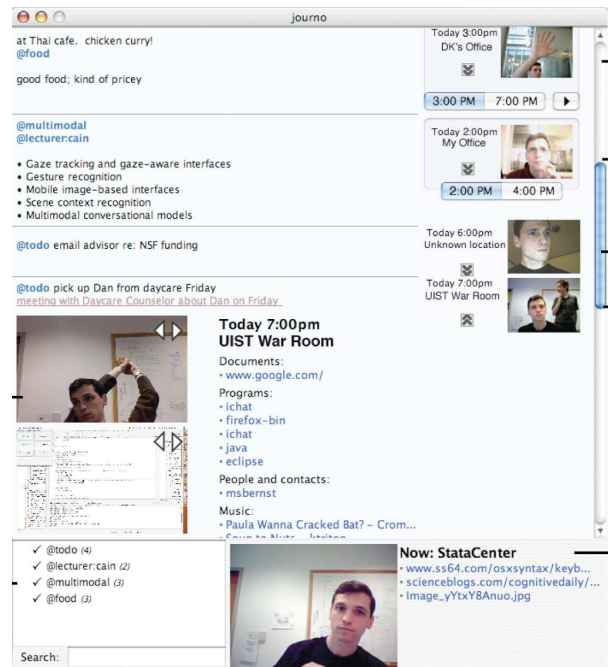


Figure 2. Jourknow, our prototype information scrap management tool.

appropriate for information scraps. Reminder scraps, for example, are intended to proactively remind rather than be indexed, searched or visualized later, and temporary storage scraps's relevance to the user decreases quickly as the scraps age.

JOURKNOW: A NEW INFORMATION SCRAP TOOL

In parallel with our ethnographic study, we have been designing an information scrap management tool (Figure 2) [2]. We briefly mention Jourknow, our information scrap client, which focuses on the following ideas:

Structure Extraction. In order to elevate ambiguous or unstructured text to searchable, sortable data, we can assist in the conversion of the raw scrap rich in implicit structure to data with explicit metadata structure. Thus, "mtg. w/ Karger @ 5" becomes reified as a calendar event in the user's calendar application and searchable as such.

Context Association. We can further assist in information scrap retrieval tasks by allowing the user to query by the situation surrounding the note capture. Jourknow automatically captures and associates information surrounding the user's situation. This data includes day and time, location hints (e.g., wifi ssid), events scheduled on the calendar, and activity traces including web pages, active applications, people the user communicated with, and pictures of the desktop and the user. Users may then use a faceted browsing interface to search for notes fulfilling specific criteria.

Mobility. We are constructing a mobile version of the Jourknow client to run on users' cellular phones. The mobile version of the client is intended to support users when away

from the computer, tailored to the information capture needs when mobile and affordances the cell phone offers.

We recently completed a weeklong deployment evaluation of the Jourknow client to help determine its strengths and weaknesses in information scrap management.

CONCLUSION

In this position paper we have raised the information retrieval problem as it affects and is affected by personal information management. We examined information scraps as a particularly interesting case of personal information in this respect, considering features which might bear on IR tasks. We have seen that importing traditional IR goals and metrics into the space of information scraps fails to account for many of the distinct qualities of scraps' encoding and usual retrieval cues. We report on our efforts to bridge this gap via our prototype system Jourknow, which attempts to make easier the capture and retrieval of information scraps.

REFERENCES

1. M. Bernstein, M. V. Kleek, D. Karger, and mc schraefel. Information scraps: How and why information eludes our personal information management tools. *In Submission to ACM Transactions on Information Systems*, 2007.
2. M. V. Kleek, M. Bernstein, D. Karger, and mc schraefel. Gui? – phooey! the case for text input. In *Proc. UIST '07*, September 2007.