Information Scraps: Eluding our Personal Information Management Tools

Michael Bernstein MSBERNST@CSAIL.MIT.EDU
Max Van Kleek EMAX@CSAIL.MIT.EDU
David R. Karger KARGER@MIT.EDU
mc schraefel MC@CSAIL.MIT.EDU
MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge MA, 02139 USA

1. Abstract

In this work we introduce research on *information scraps*, or the unorganized pieces of personal information that commonly end up in our notebooks, spare pieces of paper, free text files, or e-mail. We perform an ethnographic study investigating patterns surrounding these scraps, uncovering patterns in use, content and needs. We conclude with a description of our prototype system Jourknow, intended as a next-generation information scrap management solution.

2. Introduction

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. 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*. The term conjures up several images: notes that are written on a scrap of paper, or that have been separated from the rest of our personal information collections. Information scraps can be to-dos, notes to yourself (Lin et al., 2004), passwords, song lyrics, or a variety of other information. In this paper, we investigate the existence, use, and composition of information scraps via our own focused study of the information scrap.

3. Related Work

Perhaps the most canonical information scrap is the note to oneself, which is an emerging area for PIM research. Through a series of semi-structured interviews, Lin et al. arrived at a model of such notes' lifecycle: trigger, record, transfer or maintain and refer, complete, discard or archive (Lin et al., 2004). Others (e.g., (Hayes et al., 2003)) have investigated note practices and found salient characteristics of such notes to include transience, visibility, and mobility. They further reported what was to become something of a trend for personal notes: a strong preference for paper-based media over digital media.

4. Study Method

We conducted a study consisting of 27 semi-structured interviews and artifact examinations of participants' physical and digital information scraps, at 5 different organizations. There were 13 males and 14 females; the median and mode age range was 30-35. As our questions primarily surrounded information scrap content, organization and location, and lifecycle, we chose to focus on examining the scraps themselves rather than the retrieval process.

In order to note the varying types of information scraps, we created a method by which we would look forscraps along three axes: tool, location, and type. We first inquired after tools, and then continued with location and finally type.

5. Study Results

5.1 What Do Information Scraps Contain?

In our artifact analysis, we coded each of the 533 information scraps for its data type, then consolidated similar categories. A condensed portion of the results is reproduced below.

A compelling view of information scrap forms appears when one focuses on the least frequently occurring items. The least frequently occurring types (the tail of the distribution) comprised a significant percentage of the information stored in all the scraps; in particular, forms that occurred only once comprised 13% of all scraps; twice or less, 18% of all scraps. This result suggests that the distribution of data types contained with scraps exhibits the long-tail property of statistical distributions, where the scraps containing rarer forms cumulatively match or outnumber the small set of commonly occurring forms.

5.2 Scrap Composition

We found that text written in information scraps used extremely terse language; many scraps consisted exclusively of key words, such as lists of names of people, places or objects, and raw bits of data, such as phone numbers, addresses, passwords, and other strings. We noticed a tendency to omit the subject title or description of what the data actually represents. Several participants, when sending e-mails to themselves, intentionally left the subject field blank or wrote something general such as "note to self."

5.3 Tools

We noted 51 different tools in use across our investigation, 33 digital and 18 physical. Participants maintained a small number of main locations for capturing information scraps, supplemented with less-used auxiliary tools.

E-mail was by far the most dominant digital tool for recording information scraps (74, 26.4% of digital scraps), followed by text editor (47, 16.8%; e.g., TextEdit, Notepad, Wordpad) and word processor (27, 6.4%, e.g., Microsoft Word). In the physical world, paper notebooks (94, 37.2% of physical scraps) and Post-its (60, 23.7%) were the most popular choices.

Overall, there was an approximate parity in the number of physical (253, 47.47%) and digital (280, 52.53%) information scraps we gathered. However, a chi-squared test comparing the number of digital and physical artifacts between each participant rejects the null hypothesis ($p < 0.01$), indicating that there is some dependency between the participant and the distribution of their information scraps between digital and physical tools. As one explanation, there appears to be a bimodal distribution with most participants centered at the 50% mark and a smaller group being almost entirely digital.

6. Study Analysis

6.1 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. ADMN2 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, brainstorms, and interface designs. "Before I put anything in the computer, I like to put it on the whiteboard first," ADMN4 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, ENG3 created an information scrap system to manage a library-style checkout for his privately owned construction tools, and MAN4 maintained a complex document of contact information annotated with private notes on clients.

6.2 Desired Affordances

To help support the attributes or requirements of these various scrap classes, we propose a suite of affordances designs may need to support. These are:

Lightweight Entry. When time and effort were at a premium, the fastest tool would often win out. Even seemingly minor difficulties or annoyances with tools could have striking effects on behavior. "If it takes three clicks to get it down, it's easier to e-mail," FIN1 explained. Even when data was implicitly structureable, such as with potential calendar events, participants chose the faster, structureless route of recording a scrap: a string literal such as "mtg @ 5pm in cafe."

Freeform Contents Rather than enforcing strict encoding schemas, the most effective information scrap tools allowed the author considerable leeway in how to capture data. Data could be written, drawn, scribbled, sketched, abbreviated, annotated or rearranged however the author saw fit.

Cognitive Support Information scraps allow the author to reflect on what he or she is authoring, change some of the text, cross out ideas, or reposition elements simply to see the result. Examples of such scraps include ENG2 and ENG5's in-progress notes taken down while debugging, UI2's sketches of interface designs, and a large number of work-in-progress documents.

Visibility and Reminding Information scraps proved particularly useful to our participants when information needed to be kept visible at all times, used as a reminder, or

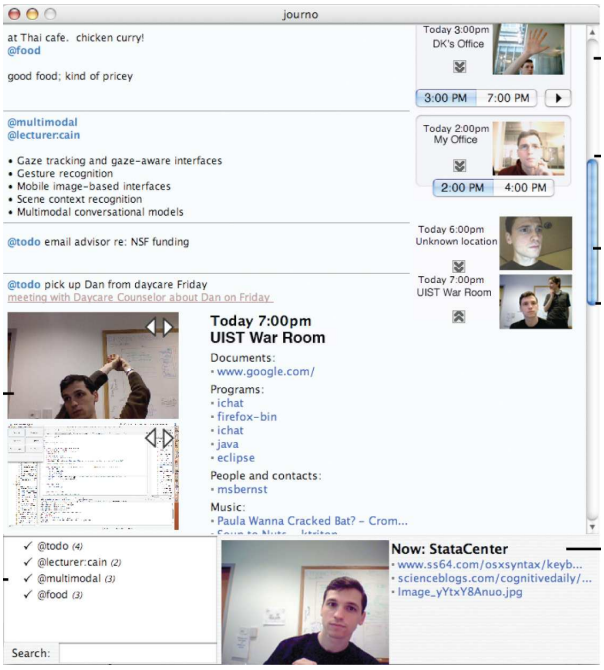

Figure 1. Jourknow User Interface

"tripped over" at an opportune moment. Physical scraps were commonly positioned in highly visible locations, and digital solutions included marking e-mails as unread and leaving files on the desktop.

Mobility When the scenario called for information workers to go mobile, participants often generated information scraps to carry important data around or to capture information as events occurred. UI2 was often mobile: she described how she would reference a note file on her smartphone with relevant conference call numbers, and was likely to send herself a voicemail.

7. Jourknow: A New Information Scrap Tool

In parallel with our ethnographic study, we have been designing an information scrap management tool (Figure 1. We briefly mention Jourknow, our information scrap client, which focuses on the following aspects:

Structure Extraction. In order to eliminate repetitive tasks, 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.

Context Association We sought for Jourknow to assist the user in recalling the situation surrounding the scrap's creation created. Jourknow therefore 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.

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. Specifically, we offer text entry via keypad, camera-phone photos and voice notation. These notes are sync'ed with the desktop Jourknow, and the desktop Jourknow's scraps are completely accessible from the mobile client.

We are currently in the process of a weeklong deployment evaluation of the Jourknow client to determine its strengths and weaknesses in information scrap management.

7.1 Conclusion

In this article we have examined the phenomenon of the information scrap, what we have informally described as personal information that is not in its eventual or proper place. We have uncovered a long tail in the type distribution of information scraps, suggesting that support for unusual or unique data types is critical to information scrap work. We have identified several desired affordances of information scrap tools: lightweight capture mechanisms, freeform and potentially unstructured data representation, cognitive support for work-in-progress manipulation, flexible organizational capacities, the ability to place scraps in the way, and mobility. These strengths drive a set of typical information scrap roles, including temporary storage, work-in-progress, reminding, archiving, and unusual data types. We are in the process of applying our insights to the design of an information scrap management tool, called Jourknow.

References

Hayes, G., Pierce, J., & Abowd, G. (2003). Practices for capturing short important thoughts. *Proc. CHI 2003* (pp. 904–905). ACM Press.

Lin, M., Lutters, W., & Kim, T. (2004). Understanding the micronote lifecycle: improving mobile support for informal notetaking. *Proc. CHI 2004* (pp. 687–694). ACM Press.