Databases – Personal Information Management

Robert Kolekta
Jared Baboo
Pheeha Machaka
Gregory Chandran

Intro

Personal information management is the storage and retrieval of all information that is used for personal use [1]. Managing personal information is a vital part of computer organization, everyday people spend time sorting out email's, files, appointments and contacts. Managing all this information can be time consuming and error prone [2], therefore systems must be developed to ease the task of managing all this information in a timely and correct manner.

Research in this field has become increasingly popular however evaluating personal information management system is not an easy task as information is collected and stored differently by different people [3].

Information is gathered by people as a natural consequence of completing other tasks [3], people collect and store information that is relevant to their personal needs. As these personal collections are unique to the user, the task of standardizing these information collections is not a trivial one [3].

Personal information management and personal knowledge management, are in most instances seen as the same field of study. However knowledge management focuses more on organization issues such as improved performance and sharing information.

Problem Description

The problem of personal information management is that one has all this information that needs to be used and managed. Each person has a specific way of accessing and storing this user specific information. Users spend significant amounts of time sorting and managing all this information. This information includes emails, appointments, contact list and any other information that is specific to a user.

A personal information management system is one that tries to handle and sort out all this information automatically. It organizes and prioritizes tasks and pieces of information that could be used by the user.
Such systems have methods and techniques so that all this stored information is readily accessible to the user, in the most basic instance the users’ needs to specify what information is needed and then the system would retrieve the relevant information. A similar system has emerged on the web in the form of a search engine. A web search engine receives a query from the user and then searches all relevant pages on the web returning results that meet the criteria specified by the user, it too sorts the information so that the most relevant information is displayed first.

Email is used vastly throughout the world as a means to communicate, sorting all this communication is a tedious task that get done daily and is time consuming. Systems are in places that allow users manage their emails and organize their tasks and calendars, however these systems are specific to email, if a user wishes to save attachments then the location of the attachment is handled completely by the user along with how the saved file is used and accessed. A good email client should satisfy three main categories [4], namely task management, personal archiving and contact information. Task management would allow users to classify emails according to tasks, this would allow users to better manage related materials. Personal archiving relates to how emails are stored and retrieved by the users and contact management relates as to how a user email contacts are sorted and accessed. A system such as MS Outlook is an example of a system that satisfies those three categories.

Bergman et al. Says that most personal information management tools have focused on specific tools and there has been little consideration as to how all these tools are used together. There is a great deal of work simply to achieve personal information management within the email section. A great challenge is to integrate all the aspects of PIM so that users can use one system to manage their emails, tasks, personal notes, address books, calendars and from music files to project files.

Recent Research in Personal Information Management
Looking at recent work involving Personal information management we discover that the research can be broken down into three categories. Firstly research can be divided into research that aims to eliminate the need for personal information management. Secondly we can break the research down into a section that focuses on understanding how people manage and interact with personal information to facilitate this. The last section this area of research can be divided into is the study of shared information management.

Eliminating Personal Information Management
Recently studies have been undertaken with an aim in mind to lessen the overhead placed on the user to manage their personal information. These studies question the need for users to actively manage and organize their information and present search functionality as a possible tool to facilitate the elimination of this overhead.

Studies on that aim to eliminate the need for personal information management present searching instead of actively managing information organization[6]. The conjecture is that the traditional
organizational methods used hierarchical organization of data that must be navigated to locate information, can:

1. Can hide information from the user

2. Can reduce the chance that a user locates the desired information

3. Forces a user to not only actively manage the placement of information in specific locations but also retain the information in their heads as to the location of this information

**Understanding Personal Information Management**

Personal information management is a relatively new research topic. This implies that we do not yet fully understand the way in which users manage their information. This section of research attempts to answer this question and pose solutions to improve and understand this.

Research in this area focuses on understanding the user and their interactions with PIM and PIM tools with aims to improve PIM tools and methodologies i.e. improve the design of PIM tools and facilitate the users interaction with personal information and PIM tools[9, 12].

**Group Information Management**

Personal information management involves user managing information that relates to them. Some research now asks what about when a user shares information or has information in which they are not the only owner contributor or manager.

Studies on Group Information Management studies personal information that intentionally or unintentionally is shared with others[10, 13]. Technology that this involves includes social networking sites, medical information, personal calendars, etc.

**Comparison and Contrast**

When looking at group information management the question of whether searching is better can be raised. With information that is not stored locally is it as efficient to search for the information being looked for than to navigate to it. In the case of sending a search over the Internet for information will the correct information be returned and then we should ask it will be returned in the state we wish it to or has it been dynamically changed. Also looking at search while this may be useful there is evidence that users might prefer to navigate to their information than searching for it. Idea’s this brings to mind is that of linking, connection and brainstorming. By navigating to information the user can link seeing different objects together to form a mental path to the information they are seeking. Also by navigating the user may be receiving different stimuli that may assist them by augmenting the information they are seeking and by seeing different information that they must navigate past they could make connections between groups of information to guide them to the information they are seeking.
Recent Research Paper

Recent Paper Research – Information Scraps, How and Why Information escapes our personal management tools.[19]

The recent research paper conducted at MIT focuses on a very interesting and significant part of our everyday lives. Personal information management (PIM), as we know it can be summarized as the practice and the observing of everyday tasks that people carry out in order to obtain, organize, maintain and retrieve and use information artifacts (digital and paper-based) for the fulfillment of their roles, socially and professionally. One of the objectives of PIM is to ensure the availability of information, whenever or wherever it is required in a complete and acceptable form. This necessity facilitated the emergence of various software tools that attempted to manage ones personal information in a way that minimizes issues like quick information access and organization of data. However even though numerous tools exist to assist one in this area, there is still a lot of information that people store outside the scope of such tools. M. Bernstein et al. refer to this as information scraps. A detailed study into the categorization and definition of these “information scraps” and the reasons for their existence is performed, to shed more light on designing better PIM tools that will cater for these categories of information.

Introduction

Information scraps can be defined as the personal information that is scribbled on stick-it notes, corners of rough sheets of paper, email messages sent to ourselves, miscellaneous digital text files created by the user to input data that could not be stored elsewhere. Such data could refer to anything from ideas, contact data, reference numbers (e.g. a tracking number for a postal package to be received), reminders (birthdays, events and meetings), meaningless note taking and writings, etc., the list is endless. Regarding PIM, these ‘free’ forms of information recording or storage despite the PIM tools available tells us that one prefers them to using a PIM tool. In some cases, the tools simply don’t exist, and other cases see people using these tools to store information that was not meant to be stored in that way.

In an attempt to discover problems and opportunities pertaining to the current available PIM tools, a study was conducted, in which 27 subjects from varying sizes of firms and professions were interviewed and observed at their workspaces. The study focused on the “information scrap” practices of the subjects and their use of PIM tools. The following questions were kept in mind:

Why do people store, scribble or keep information scraps in places other than the available PIM tools.

What function do the information scraps play in our work and other interactions?
How might we improve the designs of PIM tools to be more versatile and accommodating in terms of the various information scrap practices?

**The study of PIM**

A key factor in this area is the introduction of new PIM software that does more but ensures that the complexity of its operation and use are minimized and not increased, as is the case with a lot of PIM initiatives. The study therefore must observe users information management across various tools and over a period of time so as to get a complete picture relating to their entire PIM scope. A common mistake is to study a sole application or area of one’s PIM.

M. Bernstein et al go on to investigate the nature and use of information scraps. Types of information scraps and the tools used to collect them are observed. Information scraps, defined for the purpose of the experiment as “an information item that falls outside all PIM tools designed to manage it” (by M. Bernstein et al) can be seen as serving a specific role to its author.

These roles happen to be, temporary storage, cognitive support, reminding, archiving and the writing down of uncategorized data.

- **Cognitive support** – brainstorming, thought processes etc
- **Archiving** – Storing important information for longer periods of time, like letters, passwords, recipes.
- **Reminding** - To ensure future events and their occurrences are not missed with the use of date-wise and / or time reminders.
- **Uncategorized information** – saved and stored as the data did not fit into the users existing PIM tools.

With these roles, we can deduce a set of design principles that could form the basis of PIM tools to come. Easy entry, unrestricted content, flexibility, mobility and visibility can be good starting points.

**Data types of information scraps**

An attempt to categorize these information scraps and assess their typical lifecycle in a users information processing cycle by Lin et al. (2004) led to an observation that information scraps are triggered, recorded, transferred or maintained, completed and then archived or discarded. The format preferred by users for information scraps is paper-based. This could be due to habitual factors, the opportunity cost of the increased effort of setting up a digital entry, for example, a user would rather scribble on a piece of note paper rather than boot up a computer and run the desired PIM tool to input
a small note. The time factor in relation to the amount of data to be stored plays an important role in
the choice of digital or paper-based input choice.

- **Email usage** – email usage has been used for numerous purposes that resemble features of
  information scraps. For example, emails marked as unread, or left in the inbox as reminders, drafts
  saved as to-do lists and emails emailed to oneself for the purposes of reminders or archiving.
- **To-do lists** form a significant part of information scrap. They occur in an unorganized manner in the
  form of both paper based and digital formats. Usually these notes are written, in a short,
  abbreviated vague manner that is meaningful to the author only. Another interesting aspect is the
  way in which they are arranged and stored to ensure “visibility” as they are written or stuck to
  walls, notice boards, and even on the back of your hand
- **Calendar information scraps** – calendar tools also exhibit characteristics of information scraps by
  allowing users to keep a variety of time-based and appointment oriented events, messages to self,
  record of events occurred, to-do lists etc.
- **Camera phones**, with their multimedia messaging and interaction with online blogs have become
  popular collectors of information scraps. The ability to add pictures and photographs to everyday
  personal experiences like a scenic view or interesting memorable street poster is an enriching
  experience. However such photos can be seen as information scrap as they do not fit into any
  specific personal information categories and be lacking purpose wise.

**Organisational practice and problems**

Studies show that office workers’ document organizational habits lean towards paper based media but
neglect and avoid filing practices. Studies by Sellen and Harper [2003] found various reasons for the
preference of paper in the office space. Amongst these were the ease of annotation, navigation, spatial
reorganization and the ability to collaborate. These factors align with many information scrap needs. For
example, people usually combine unrelated information onto on sticky-notes like serial numbers,
contact details and a birthday reminder, thus amalgamating unrelated information.

The problem of “ephemeral information” termed by Barreau and Nardi refers to the situation where
information is either poorly filed or not at all. This results in the accumulation of larger volumes of
unorganized data that becomes difficult to manage. In comparison, the lack of need or motivation to
organize less important data is applicable to information scraps as well.

Another such problem regarding PIM is information fragmentation, which involves information
storage across multiple platforms on a particular single topic. This leads to the difficulty of synthesizing
information from multiple sources about a particular topic. It can therefore be said that due to the
minimal effort, haphazard-data capture nature of information scraps, they are susceptible to such
fragmentation.
Research Goals & method

M. Bernstein et al. proceed to discuss research goals using questions like “why do information scraps exist?”, “What kind of information do they hold?”, “why they end up in the tool of medium they do” and “how they evolve through their lifetime?”. These questions provide the foundation for the method and results of the study.

Information scrap is then categorized into three different sets, namely Type variety, structure and expression and tools which encompass the kind and variety of data that will be encoded, the comparison between the information scrap and its digital version and the adaptation of tools to hold information previously not catered for.

For the bulk of the study, 27 interviews were conducted, examining the participants paper-based and digital information scraps. The three tools mentioned above were the key elements in identifying information scraps. Below is a table of results as compiled by M. Bernstein et al. showing the categorization of the various information scraps.

<table>
<thead>
<tr>
<th>Triangulation Perspectives</th>
<th>Examples</th>
<th>What was targeted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tools</td>
<td>Email</td>
<td>Messages that do not serve communication purposes: emails sent to oneself, in the Drafts folder, or archived in the inbox.</td>
</tr>
<tr>
<td>Calendar</td>
<td>Calendar entries that did not correspond to actual events; use of the details field.</td>
<td></td>
</tr>
<tr>
<td>Bookmarks</td>
<td>Bookmarks carrying information beyond just a pointer to a web page—for example, “todo” or “thread” bookmark folders.</td>
<td></td>
</tr>
<tr>
<td>Physical Notebooks</td>
<td>All available data (this location commonly holds information scraps).</td>
<td></td>
</tr>
<tr>
<td>Physical Post-it Notes</td>
<td>All available data (this location commonly holds information scraps).</td>
<td></td>
</tr>
<tr>
<td>Notetaking Applications</td>
<td>All available data (this location commonly holds information scraps).</td>
<td></td>
</tr>
<tr>
<td>Freeware text files</td>
<td>“todo.txt” or “todo.doe” files containing personal notes, to-dos, and other data.</td>
<td></td>
</tr>
<tr>
<td>Locations</td>
<td>Computer Desktop</td>
<td>Documents of short-term interest and notes to self.</td>
</tr>
<tr>
<td>Physical Desktop</td>
<td>Freeware notes and documents of short-term interest.</td>
<td></td>
</tr>
<tr>
<td>“Miscellaneous” Folder</td>
<td>Data that was difficult to categorize.</td>
<td></td>
</tr>
<tr>
<td>Office wall and whiteboard</td>
<td>Participant-authored decorations or annotations.</td>
<td></td>
</tr>
<tr>
<td>Types</td>
<td>Reminders and To-dos</td>
<td>To-dos not in the todo manager or that did not fit the to-do manager’s schema.</td>
</tr>
<tr>
<td>How-to guides</td>
<td>All examples (the known application to organize this information).</td>
<td></td>
</tr>
<tr>
<td>URLs of interest or quotes from web sites</td>
<td>Examples not held in a bookmarking utility.</td>
<td></td>
</tr>
<tr>
<td>Contact information</td>
<td>Examples not held in a contact utility.</td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td>All examples (common information scrap).</td>
<td></td>
</tr>
<tr>
<td>Short pieces of data (e.g., phone numbers, passwords, serial numbers, thank-you note lists)</td>
<td>All examples (common information scrap).</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 - M. Bernstein et al.
**Results**

From the results compiled by M. Bernstein et al. (table below) we can see that information scraps were dominant in the form of paper notebooks, e-mail and post-it notes. The trend we can observe is that few extremely popular tools were used followed by a large number of less frequently used techniques.

![Figure 2 - Results compiled by M. Bernstein et al.](image-url)

Upon analysis and interpretation of the study, the following design principles were extracted from the study that with a view of improving and further ensuring the inclusion of information scraps into the PIM tools.

**Lightweight Capture** – Allow the input of information ensuring the least effort and distraction

**Mobility and Availability** – multiple or a variety of data or information capture methods will be necessary in different scenarios, this should be catered for, by making the PIM tool more versatile.

**Visibility and Reminding** – Information should be displayed appropriately, in the right format, in the right place, taking into consideration the user centered design heuristics, as well as a timely display avoiding the traditional problems of information scraps clutter.
Flexible Content – Information should be able to be recorded regardless of its data type or format. The information should also be able to be merged with other categories of information, allowing for the ease of information integration as the user or author sees fit.

Flexible Schema – data may not adhere to existing PIM molds and this should be catered for.

Flexible Categories – Data and the way it is organized should be able to be transformed or altered from an organizational point of view. The conventional rigidity of PIM tools should be avoided. Ordering and structured storage of information should be applicable to all unorganized data.

Flexible Linkage – The flow of related information should be preserved and all data should be, at the user’s request be able to be seamlessly linked, allowing the user a sound collection of information in terms of data links and related stories or topics.

The above design principles have emerged from this study to further and advance the current technology and operation of PIM tools to better encapsulate information from our everyday lives. This way our PIM tools can take one step closer to eliminating information scraps and their redundancies in terms of organization and efficient storage and access.

Future Work

The Future of E-Mail
In the field of Personal Information Management, the Email has been the major focus. Today, people engage in a variety of complex behaviors using email, such as project management, collaboration, meeting scheduling, to-do tracking and other activities [15].

Deciding Action
Email consumes a lot of time and attention in the workplace [18]. Thus, people tend to rate message according to importance (work related); message characteristics; and characteristics of message senders and receivers. In that way, a large portion of Email traffic is information that does not necessary demand a reply, and some never gets a reply [18]. Deciding on the action of an email requires the user’s insight into the meaning and importance of the email.

Recent research on email has turned into content-based techniques that identify task-related aspects of messages, so that people can easily act upon receiving the messages. There has been research that focuses on summarizing messages so that it is easier to extract key information [16].
There is also research into identifying the relationship between the message and the activities the user is currently engaged.

Future developments suggest that, instead of distributing such information through email, rather send it through Instant Messaging. Muller, Geyer, Brownholtz from IBM, are working on a unified communication application, blurring the boundaries between email chat and shared workspaces and offering integration of six types of objects: message, chat, file, folder, annotated screenshots and to-do items.

**Stuff I’ve Seen**

The user’s work involves finding and re-using previously seen information; the information may come from an email, document, web page, appointment and others [17]. There has been work done by Microsoft that enables users to remember and discover information. A system was developed, Stuff I’ve Seen (SIS), which facilitates information re-use. The user need not organize their data into folders or hierarchies, the system allows the user, to use any property (date, file type and author) or keyword, to find what they are looking for. The search uses two capabilities: first, it cuts across the many disparate sources of information we encounter daily (from web browsers, to email); and second, it cannot be restricted to keyword search but include other kinds of information associated with the item or context in which it was encountered (its metadata) [17]. The system’s rich search capabilities can be used whether or not an item is explicitly saved in a folder. The system takes advantage of the user’s context, to pro-actively find task-relevant information. It will analyze the current context, identify important words or metadata and automatically generate queries to find related information.

There is great research into new search capabilities that will change the way people organize information; this will make explicit filing and organizing far less important for retrieving personal information.

**MyLifeBits**

Microsoft is currently running a project, MyLifeBits. This project not only keeps the user’s personal information like documents, emails and appointments, it will also allow multimedia sounds, pictures and video. This was developed to fulfill the Memex vision first posited by Vannevar Bush in 1945.

The MyLifeBits project is an effort to implement a personal digital store. It is Memex, extended beyond Bush’s vision to handle audio and video, to perform database style queries, and to allow multiple visualizations in the user interface.

In this system, the user is not limited to organizing the data in a hierarchy or folders. Some items may belong to more than one category, making it difficult to arrange. The system allows the user to assign an object to zero more than one set or a collection; using links. This allows hierarchy but does not enforce it [14]
The system also allows the user to annotate non-text media like, pictures, audio and video. In this way, the media is given more meaning and value; so that the user can easily remember what it is; what was their thought about it; what was someone else’s thought about it. This will allow the user to construct a story about the media, making it very valuable and easy to recall for them.

Links and annotations make it easier for the user to recall the context in which the media was taken, and makes it easy for the search function to find the media easily.
Summary

In concluding, we can see that Personal Information Management and their tools need to be further developed to improve the management of information. As information, especially personal information is stored and accessed differently by each person; it is therefore a difficult task to standardize tools to deal with this varied unique information.

However the need for better information management tools continues to exist and transform as more and more time is taken up in the organization information.

Email, one of the most universally used communication methods, and its PIM tool, the email client are discussed and analyzed. The key aspects like task management, personal archiving and contact info management are identified as the most important attributes to a successful email application.

A common trend in PIM tools, their use and focus on specific applications without the need to encompass a larger category of information and integrate multiple information management tools is discussed. Recent research into PIM falls into the categories, “the need to eliminate PIM”, “Understand how people manage and interact with information” and “shared information management”. The first discusses reducing the load and stress on a user to manage his PIM, using search functions. The second studies the interaction between users and PIM tools and the last looks at information that is shared with others where multiple authors contribute to the data and concepts such as social networking.

A recent research paper based on “information scraps” and why they elude our PIM tools is reviewed. It is a study of one’s everyday lives and information that we scribble down or have no desire to input into a PIM tool. Current problems with information tools were identified through a study of professionals from different firms and their information scrap practices. Information scraps and their roles like cognitive support, archiving, reminding were discussed. Results from the study provided new design principles like lightweight capture, mobility and availability of data and the visibility and reminding aspects of information which proved to be future key points.

Future work with PIM and their tools focus on Email as the cornerstone, expanding into a multi-faceted form of interaction involving, project management, meeting scheduling, to-do tracking etc. As email consumes time, an important innovation explained was that of content summarization and importance determination of future email applications.

IT organizations are undertaking research on finally integrating PIM tools across information platforms, unifying the process of information management. IBM is performing work to integrate email using instant messaging, to encompass all aspects like messages, chat, files, folders, to-do items, etc. Microsoft is working on re-use of information and more powerful search methods to retrieve the data a user requires. This is different in the sense that the system performs a context-related search across multiple sources of information. MyLifeBits, an online digital store, to collect a users’ information and organize it into one place is another Microsoft initiative. A flexible organization structure is emphasized, with the ability to store videos, pictures and documents.
Bibliography


[10] Erickson, T. From PIM to GIM: personal information management in group contexts.


Questions
1. Why do we need personal information systems
   a) To help communicate with other people
   b) To help in setting up appointments
   c) To reduce the time and effort taken in organizing ones information
   d) To help people sort out their emails

2. What is an advantage does navigating to find information in PIM has over searching to find it?
   a.) Information can be found faster by navigating.
   b.) Navigating allows people to view related information
   c.) People are find more accurate results if they navigate
   d.) Navigating is better as people prefer it

3. Who is MyLifeBits' experimental subject?
   a) Gordon Bell
   b) Bill Gates
   c) Steve Balmer
   d) Craig Mundie

4. What is Information Scrap
   a) Contacts in an Address book
   b) Email in an inbox
   c) Email sent to one self as a reminder
   d) A PDA to-do list containing to do items