doi:10.1093/iwcomp/xxxxxx

Transference of PIM Research Prototype Concepts to the Mainstream: successes or failures

Personal Information Management (PIM) refers to the practice and the study of how people acquire, organize, maintain, retrieve, archive and discard information for various reasons in physical and digital worlds. Many PIM tools are available for managing information on our desktop computers while many research prototypes tried to augment or replace them. The development of these tools was based on knowledge drawn form the fields of psychology, human computer interaction, information retrieval, knowledge management and research in the PIM field. Different metaphors and ways of organizing were introduced. However, the prevailing beliefs are that most of these prototypes were not extensively tested and that the radical design (not addressing real world issues) and quick abandonment of prototypes prevented transfer to mainstream products. This paper looks at what has been developed and learnt, what has been transferred to mainstream applications, discusses the possible reasons behind these trends and challenges some parts of the above mentioned beliefs.

Categories and subject descriptors: Critical Reviews

Keywords: Personal Information Management, Information Storage, Information visualisation, Information retrieval, Document management

Responsible Editorial Board Member: Name

RESEARCH HIGHLIGHTS

• Survey of Personal Information Management research prototypes.
• Discussion on how these prototypes were evaluated and how the research focus has changed during the last 20 years.
• Overview of what technologies have been transferred into publicly available software and possible reasons why this occurred.

1. INTRODUCTION

Researchers have long studied how people manage different types of documents and several problems have been found in both physical and digital domains. Many studies have been carried out on how people manage files (Barreau and Nardi, 1995; Ravasio et al., 2004), email (Whittaker and Sidner, 1996) and web bookmarks (Abrams and Baecker, 1997) as separate entities. It is not surprising that PIM prototypes tried to overcome problems people had when managing different types of information held by disparate systems. In the last decade, when much information became digital and with the advance of portable devices, a problem of how information fragmentation affects task performance has become prominent (Boardman et al., 2003; Bergman et al., 2006; Tungare and Pérez-Quiñones, 2009). Many prototypes tried to unify the information space on personal computers and beyond. Focus also shifted to research on how people manage their information together in physical and digital domains (Jones et al., 2002; Bondarenko and Janssen, 2005), as well as how they manage information on different devices and across different tools on these devices (Tungare and Pérez-Quiñones, 2009). Recently managing personal information on, from and with the help of the web has also been included in studies (Moore et al., 2010).

Some authors have already classified a small set of PIM prototypes. Boardman classified research prototypes in four groups: (a) tools that improve management of specific information type (files, email, web bookmarks) where integration of other information types is not a goal, (b) tools that try to improve integration between distinctive tools, (c) tools that are focused on managing one information type but embed additional support for managing other information types as well, and (d) tools to manage different information types within a new, unified design (Boardman, 2004). Karger's classification focused on information integration from the point of underlying technology (Karger, 2007; Karger and Jones, 2006). He divided PIM tools in three main groups: (a) visual unification, (b) unification based on standard data operations like copying and pasting text from one application to another and (c) unification based on metadata (different groupings of information items, cross referencing between information, attributes and relations).

This paper does not focus on integration alone and describes more than 100 prototypes divided by information type. Three sections describe prototypes that addressed problems in managing each information type separately: emails (3), files (4) and web bookmarks and history (5). This is followed by a section dedicated to prototypes that integrate different information types to overcome the problems of managing semantically related information (6). In each section, prototypes are grouped in subsections and compared by the functionalities they implemented to overcome similar drawbacks. They are also listed in time order whenever possible (with a year in brackets following the name) to show the course of development. Each subsection ends with a short discussion about what functionality is used in (or has been transferred to) mainstream applications. The paper concludes with an overview discussion (7) and a conclusion (8) about two beliefs: that prototypes are not extensively tested (Bergman et al., 2004; Boardman, 2004) and that transference of PIM prototypes to the mainstream is low (Bergman et al., 2004; Whittaker et al., 2000). These beliefs arose from early works and this paper wishes to discuss if they are still valid.

2. METHOD

Despite the adoption of mobile devices and web based PIM services, most PIM still occurs on or is synchronised with personal computers (PC) (Jones et al., 2008; Weinreich et al., 2008; Whittaker et al., 2011). Since the advent of PCs, many PIM research prototypes were developed drawing from the fields of artificial intelligence, information retrieval, human-computer interaction and others. The categorisation in this paper includes 146 desktop PIM research prototypes supporting management and visualisation of the three most common information types found on personal computers: files or documents, email and web bookmarks/history.

The list of examined prototypes started from our own paper collection and was expanded with references from these papers, IEEE Xplore, ACM digital library and Google scholar searches. We also created a web site dedicated to PIM prototypes (X) in August 2010, and invited others to add more to the list (the web page is listed in the X article and attracts approximately 150 users weekly). The list is not exhaustive, but the large number of prototypes covered is significant enough for spotting research trends. There are a vast number of other public information management prototypes, prototypes unifying physical and digital information spaces, free form information management and temporal desktop unification prototypes (see X); however we have not taken such tools into account. Our aim was to cover personal information on personal computers and its persistent information structures that draw sufficient attention from the research community with a sensible and sufficient numbers of prototypes.

We organized all prototypes by the main information type they supported (email, files, web bookmarks and history, and information integration). Under each type we additionally grouped prototypes based on what problems these prototypes tried to solve, how these problems were addressed or solutions proposed. Each section is followed by a short discussion paragraph of what has been transferred to the mainstream applications. This was based on our own tracking of development in the industry and discussions on different sections with other colleagues and friends. Through this process, we were able to spot trends of what is used in everyday PIM applications – one of the main aims of our work.

Another aim was to look at whether there is a connection between the fact that a prototype has been extensively evaluated and its appearance on the market. To find out how much each prototype was evaluated, we used papers where prototypes were described as a source. If the paper of a particular prototype did not mention or describe a user study, we used Google scholar to find other papers referencing it. In particular we searched for papers of the same author(s) to see if the discussed prototype was evaluated in some of their later works. In addition we searched authors' own web sites and their lists of publications as well as digital libraries to see whether any other paper discussed a user study regarding the prototype. Some prototypes were not evaluated as a whole and only bits of them were extensively tested by users, while others focused on algorithms and how well they performed compared to other algorithms. All these findings are presented in the discussion section at the end after the description of prototypes by information type.

3. EMAIL AND OTHER COMMUNICATION TOOLS

Email has become one of the most popular asynchronous communication tool and computer application. Most of us spend a significant amount of time using the email client checking, reading and replying to email messages (Ducheneaut and Bellotti, 2001; Mackay, 1988). Sometimes even a quarter of a working day's tasks include email (Ducheneaut and Bellotti, 2001). One of the problems people face with email is email overload. Concerns about information overload had already been expressed in the 80's (Denning, 1982) and it has later been studied by several researchers (Mackay, 1988; Whittaker and Sidner, 1996; Fisher et al., 2006). Below we divided email prototypes in five groups which sometime overlap: prototypes that (1) provide some sort of automation or help with rules and filters and focus on organization, (2) provide context clues for easing retrieval, (3) support tasks and have other information types embedded, and (4) provide a different visualisation or perspective of the inbox.

3.1. Rules, filters, agents and automatic classification

Information Lens (1987) addressed information overload with rules and filters to automatically file emails, and tried spotting possibly relevant information to automatically forward it to others (Malone et al., 1987). **Tapestry (1992)** introduced user and collaborative based email filtering to help users organize email (Goldberg et al., 1992). Email client **Maxims (1994)** learned from users' actions and it could prioritize, delete and archive messages on behalf of users (if its prediction confidence level was above a threshold level) or it suggested possible actions (Maes, 1994). Automatic classification was also used in email client **Re:Agent (1998)** with text mining based learning capabilities (Boone, 1998). **CAFE (1998)** introduced three different modes of automatic email management based on the time users have at hand: busy, cool, and curious mode (Takkinen and Shahmehri, 1998). Segal and Kephart suggested that rules and filters should be applied only after emails are read (Segal and Kephart, 1999). Their email client **Mailcat (1999)** (known later as SwiftFile (Segal and Kephart, 2000)) processed all incoming email though an adaptive classifier and proposed the three most probable folders to which emails could be filed. Another email parsing technique (not related to email organisation) was used in the **Apple Data Detectors (1998)** that searched for structured information scraps in emails (e.g. dates, postal and email addresses, phone numbers) and assisted transfer to other relevant PIM applications (e.g. calendar or address book) (Nardi et al., 1998).

Usage in practice

Rules and filters are present in almost all email clients, although, to our knowledge, it is not known to what extent these are used. One potential problem with rules and filters is that users do not like them in general (Whittaker and Sidner, 1996; Bellotti et al., 2003). Automatically filed information can easily be missed and automatically sent information deprives users of control over what is sent out. It is also not known to what extent users would tolerate errors of automatic classification. Automation is usually used for detecting spam; however a Google Mail feature "Priority inbox" incorporates some of the above ideas and learns and sorts email based on users' actions such as: which email is read, replied to, the sender, etc. Data Detectors are also used in Apple's PIM suite (iCal, Mail and Address book). Mailcat is an extension of Lotus Notes under the name Swiftfile. Some email clients (such as Thunderbird) parse email text and detect if a user wants to send an attachment while some clients (e.g. Thunderbird, Gmail) allow one click "archiving" from the inbox. Agents and automation are being implemented in mainstream email clients, albeit very slowly (many years after email has started to be commercially used in the 1980s or started to be widely adopted in the middle of 1990s).

3.2. Contextual clues for retrieving and reminding

The context in which information is viewed, classified and remembered was found to be very important for later access and retrieval (Tulving and Pearlstone, 1966; Kwasnik, 1989). Several prototypes assist retrieval by providing relevant information, different views of the inbox and conversations. Email client **Mona (1993)** could show the context of each conversation (email thread) graphically (Cockburn and Thimbleby, 1993). Similarly the **Thread-based email client (2001)** joined emails and replies in threads which was positively accepted in the user study (Venolia et al., 2001).

It was proven that category interfaces help users find relevant information on the web quicker than with plain list interfaces (Dumais et al., 2001). **BiFrost (2002)** organized email in the inbox by categories which changed based on the context: email linked to a current day and calendar inputs, email from important people, email replies, email with more than one recipient, etc (Bälter and Sidner, 2002). **Sudarsky's prototype (2002)** allowed users to browse email ordered in hierarchic folders by attributes such as (1) domain names with top domains (.com, .net, .org, etc.) as the first hierarchy level, (2) company's or organization's names retrieved from domain names and (3) based on senders last name (Sudarsky and Hjelsvold, 2002). Email of a selected folder was presented as squares, organized in a grid by time and sender (based on TimeSpace mentioned in section 3.3) or in time-ordered threads. The search emphasised relevant email with colour, email content, sender's photo and number of recipients.

Zest (2002) extracted initial questions and replies and constructed threads based on content as opposed to whole emails (Yee, 2002). **Threadmap (2005)** used a treemap algorithm (Johnson and Shneiderman, 1991) to display and help identify positions, arguments, and evidences of conversations (using a special conversation convention) (Yee and Hearst, 2005). It showed only a few levels of a thread (real estate limits) but allowed users to dig down. Zest and Threadmaps were meant for large scale conversations (e.g. newsgroups); however, they could be used for visualising threads in the inbox as well.

EzMail (2004) provided (1) visualisations for individual messages, their properties and user made annotations, and (2) visualisations of messages as components of threads, providing contextual information of conversations (Samiei et al., 2004). **ReMail (2004)** (successor of Extreme Blue and Dremail) was equipped with even more contextual clues such as a time scale, a graphical representation of email threads (time ordered nodes of email connected by arcs called Thread Maps or Thread Arcs (Kerr, 2003)), list of people involved in a conversation, annotations, discovery of dates and time, showing relationships between messages in the same folder (called Message Maps), annotations and selective displays (Kerr and Wilcox, 2004; Rohall et al., 2001; Rohall and Gruen, 2002). **Bluemail (2008)** combined threading, tagging, folders and search in the same client. It uses threads as a fundamental organization unit and actions can be preformed on threads instead of individual emails. (Tang et al., 2008)

Usage in practice

The concepts such as time scale (e.g. Thunderbird's search results) and email threads (e.g. Gmail, Outlook, Thunderbird's "Gmail Conversations" extension) are implemented in many available email clients. Even visualisation of threads is available (e.g. Thunderbird's extension ThreadVis visualises threads in a similar way to Thread Arcs). Several clients also support easy one-click annotating (flagging, tagging, labelling) which (besides multi-categorizing email) provides contextual clues for easier retrieval. Some even have rules for automatic tagging (Gmail) which does not move emails out of sight, but rather give clues to what might be important (e.g. tag all emails from a certain domain) which, to some extent, allows grouping email in the inbox.

Other research prototypes mentioned in the next two subsections also provide several clues for reminding and retrieving email, but their focus was on supporting tasks or visualising hidden patterns in email archives.

3.3. Task support

Email has always been extensively used for other activities besides its primary role - communication (Mackay, 1988; Ducheneaut and Bellotti, 2001). People manage tasks (projects, activities), delegate work, backup files, send themselves reminders and use email for several other activities which were not envisioned by email creators. This flexibility of email has its drawbacks as well, as important emails can easily be buried in the inbox with less important ones. The diversity of usage has led to several prototypes to support it – one such usage is task/project management.

Raton Laveur (2000) integrated calendar, to-do list and notes in a regular inbox to provide support for task reminders (Bellotti and Smith, 2000). To reduce clutter (of mixed information in the inbox) and cognitive effort (to file information), it allowed grouping and searching for information with users' defined and predefined queries (unseen, completed to-dos, etc.). **TaskMaster (2003)** also had task management incorporated in the email client where each task could include a list of emails, URL's, attached files, associated clusters of actions and a graphically presented due date (Bellotti et al., 2003). Another task based email prototype is **SmartOffice (2012)** which provides semi-automatic classification of email based on activities. It also displays files relevant to the email at hand (Lampasona et al., 2012).

TimeStore (1997) abandoned the traditional inbox and used a time-based email visualisation (Yiu et al., 1997). Emails were presented as dots on a two dimensional grid with time on a horizontal axis and contacts on a vertical axis. Contacts served as social reminders and showed rhythms of conversations, while dots could be annotated as tasks and were visually emphasised if moved too far away in the past. The tool inspired another grid based email prototype called **TaskView (2002)** with dates on the top axis and email subjects on a side (day/subject as opposed to day/sender in TimeStore) (Gwizdka, 2002). While both prototypes provide a better overview of tasks in emails, they also lacked the regular inbox view in which users can search for their emails by subjects, dates, senders and other standard email attributes.

Usage in practice

Major email clients have task management embedded. Outlook 2007 has a task list, to which emails can be dragged to as new tasks. Gmail has an extension called Taskforce which splits emails in three categories: information, action and broadcast (action vs. information orientation of information types has been discussed in (Whittaker, 2011)). Tasks can be created from its interface or from email, can be shared and can have files attached as well. Thunderbird also has several extensions that provide reminding capabilities of to-do's and tasks integrated in a calendar. Further integration for task/project support in everyday application is unknown to authors. With current integration possibilities people save email messages as files in a file hierarchy or print them, although such information looses the context of its conversation thread (Bondarenko and Janssen, 2005). Files can also be added to email as attachments, however this does not provide any serious (task or project) related information management. Nevertheless, it has been observed that email, file and bookmarks hierarchies overlap especially where information is managed by projects. Even if in one study users did not wish to manage information together in a provided PIM prototype because of a different acquisition type of each information type (e.g. incremental file hierarchy growth versus lack of control over email flow) (Boardman et al., 2003), other studies reported on positive reactions on presented information integration (Bergman et al., 2008). Prototypes that integrate more information types are discussed in section 6.

3.4. Different visualisations of the inbox

The idea behind most of the visualisations presented in this section is to provide a view to missing context by revealing habits, relationships and other data and patterns that are hidden within the email archives. Most of these prototypes are not focused on email management at all and would only work as an add on to present email applications. Nonetheless they allow users to reflect upon their PIM past. We divided these prototypes based on the focus of each visualisation: time, threads, contacts, metaphor based visualisations and art visualisations. As said before, some of these prototypes belong to more than one category (eg. a mix of a social network evolving through time or threads on a time scale) and will be mentioned accordingly.

3.4.1. Time-based visualisation

TimeStore and Sudarsky's prototype (3.2) with contacts on a vertical axis revealed the rhythms of conversations with each contact. TaskView (3.3) showed timely ordered email in threads with email subjects on a vertical axis. Time ordered email threads with coloured emails as nodes in a graph were used in a predecessor of reMail (Rohall et al., 2001). This prototype used other time based visualisations such as reduced resolution email overviews on a calendar like interface with dates on the vertical axis on the top and columns of emails underneath.

Some prototypes did not even show email. **PostHistory (2004)** displayed (1) the intensity of email exchanges (depicted as squares) over time in a calendar (with size of squares revealing the amount and color the directness of email) and (2) a contacts map with the user's name in the centre and names of other contacts placed around it – the more email they exchanged, the closer they were to the user (Viégas et al., 2004). **FaMailiar (2004)** also revealed communication rhythms, patterns and statistics displaying email on a time plot (shown as different shapes based on the directness and contact's intimacy with user) (Mandic and Kerne, 2004). Similarly, a visualisation from EmailViz showed **"rhythm of a relationship" (2005)** or how email activity (beginning of a relationship, intensity and death) with each contact changed over time (Perer et al., 2006). **Mailview (2005)** also plotted emails chronologically enabling users to analyse and observe conversations' trends over time and perceive emails with similar features in a filtering and zoomable interface (Frau et al., 2005).

The Mountain (2005) visualised an email archive with layers ordered by time on a time scale, where each layer represented a different person and the thickness of each layer referred to how recently the person has been in contact with a user. It could reveal life events such as switching jobs, retiring or graduating (Viegas, 2005). **Themail (2006)** parsed content of email, extracted and displayed keywords on a screen in time ordered columns, coloured and sized based on the frequency of use and conversation they belonged to. It revealed topics of conversations with each contact over time (Viégas et al., 2006).

3.4.2. Contacts-based visualisation

Four of the above mentioned prototypes (TimeStore, contacts panel of PostHistory, "rhythm of a relationship" and the Mountain) used contacts in relation to time, where time was the component that drove visualisation. In this section the focus is on contacts. Most of the presented prototypes tried to visualise a network of connections between people found in email archive(s). There are two main approaches: (1) a user perspective on a personal social network (built either automatically or manually) or (2) the whole network visualised with no central spot. We will describe prototypes with the former approach.

Personal Map (2003) draws a social network with a user in the centre, surrounded by groups of contacts based on their co-appearance in emails and radially arranged based on the quantity of exchanged email – the more email exchanged, the nearer to the center (Farnham et al., 2003). In a similar fashion, **Social Network Fragments**

(2004) arranged contacts taken from co-occurrence of addresses in email (Viégas et al., 2004). But it added a time dimension, so one could watch the evolution of the network.

On the contrary, **ContactMap (2004)** allowed manual management of thumbnail images of contacts on a spatial area (implemented based on user preferences in a series of development circles) (Whittaker et al., 2004). Newly received emails were shown on the thumbnail images. Emails could also be sent by selecting the appropriate option on a desired recipient's image. This social desktop provided the support for two distinctive tasks: (1) reminding about social commitments (similar objective of the TimeStore) and (2) social data mining for expertise or advice.

Another set of visualisations from **EmailViz (2006)** used treemap, scatter plots and histogram timelines to reveal patterns in communication (Perer and Smith, 2006). Treemaps presented a hierarchically ordered set of boxes within boxes based on email domain. Correspondent Crowds scatter plots revealed the dynamics of communication and were generated based on the number of messages sent to the the contact against the number of messages received from the contact. The third visualisation was a histogram called AuthorLines that revealed the difference between the number of messages in conversation threads authored by the user and the number of replies initiated by others on a yearly basis.

3.4.3. Metaphor-based visualisation

Some prototypes tried to introduce different metaphors to email management. **Email nodes (2004)** let users manage their email in piles (threads, topics, projects) on a spatial zoomable environment where zooming closer to nodes or email revealed more content (Diep and Jacob, 2004). Piles were also used in **Mailstacker (2005)** (Lam, 2005). The main drawback is scalability of such systems as with vast amounts of daily received emails the workspace could easily get cluttered.

A natural metaphor of bacteria like creatures was used in **Anymails (2007)**, where contacts were categorized in 6 predefined categories, and email from each category was represented by a different animal species floating on the spatial area (Horn, 2007). The size and opacity of animals visualised age, while speed and hairiness showed its status (from hairy and speedy unread email to almost static non hairy replied to email). This prototype could reveal density of received email by groups in a selected time period. In the **Mail Garden (2007)** the inbox was arranged horizontally by time and email visualised as trees. The height of a tree reflected the length of the email it represented. Like Anymails, it had a time line on the bottom of the interface where users could select a time period (Wilkens, 2007). **Magnet Mail (2009)** used a natural metaphor of magnetism for searching (Castro and Lopes, 2009). Searching keywords were represented as magnets on a zoomable spatial area, and emails were depicted at being attracted/repulsed from the keywords.

Usage in practice

While different visualisations or views were positively accepted by participants in the studies, users usually did not want to abandon the inbox view (especially with prototypes that still allowed some email management such as ContactMap, Time Store, Time View). Such visualisations are often seen as a supplement to existing email management for easier navigation, search and revelation of some patterns.

Some attempts were made to bring social networks to email. Personal Map was partially integrated with Microsoft Outlook and enabled users to compose email to groups of people. One of the problems reported was the dynamically changing network as names appeared in different places at different times – this is fine for exploring an interesting pattern, but for serious use such dynamics seem annoying. Another tool providing social network and people search for Outlook is Xobni. Google tried to integrate social networking (Google buzz) with email. There are many aggregators that try to bring several web 2.0 services together. One of them was Vodafone 360 (discontinued) that brought together all contacts from phone, email, instant messenger and social network accounts and aggregated social network feeds. Similar features are available with Smartr for smart phones and Gmail.

Besides social networks, there are other available views of one's email archive. Gmail has several add-ons such as "Graph your inbox" which generates some statistics of usage. There are even more interesting approaches. The 3D mailbox turns email management into a game within a 3D digital world where emails are visualised as people and aircrafts (`http://www.3dmailbox.com/`) while 0boxer turns cleaning Gmail's inbox into a social competition.

4. FILES

Several studies have been carried out to better understand how users manage their files (Ravasio et al., 2004; Barreau and Nardi, 1995; Barreau, 2008). In these studies, hierarchies used to manage files were criticized for various reasons: files can only be at one place at a time, organizational structure cannot easily change over time, dynamically organized collections (according to projects, time, content, etc.) are not possible, automation in organization and archiving is missing, naming and categorizing requires high cognitive efforts, versioning not supported, and file attributes (size, time, name) are more helpful to computers than users. All these issues have lead to several research prototypes.

Some of these issues have already been addressed in mainstream file managers. Features such as aliases and virtual folders have existed for a while now. Based on estimations, aliases are used in 16-20% of the cases for retrieving information, and virtual folders in about 1% (Bergman et al., 2008). Other features such as automatic stacks or arrangement views are fairly new and it is not known yet how they affect information management.

Some researchers argue that most of the practices users employ and problems they have are related to the technology and artefacts that underlie the managing activities (Fertig et al., 1996). The static nature of hierarchies has also its supporters, who argue that it makes navigation easier (Boardman and Sasse, 2004), which is what people prefer for retrieving personal and also publicly available information (Barreau and Nardi, 1995; Bergman et al., 2008). The discussion in this section will focus on prototypes for file management only. Several cross tool prototypes also tried to solve these problems, but are presented in section 6 as they are also addressing information fragmentation.

4.1. Attributes and multi-categorization

A few prototypes tried to address rigidity of hierarchies and user unfriendly attributes. **Semantic file system or SFS (1991)** was an extension to the tree based hierarchy and provided access to files via virtual folders (Gifford et al., 1991). This folder corresponded to associative queries over an automatically indexed set of attributes. **DomainView (1999)** introduced dynamic structures based on (computer and user defined) attributes where categories were dynamically formed by search queries or manually defined by users, could overlap and be nested (Baeza-Yates et al., 1996). **Linking File System (2005)** also allowed files to have user or application specified attributes, and attributed links between files (Ames et al., 2005). Similar hierarchy/tag navigation was also supported in **TagStore (2010)** (Voit et al., 2011) and **Attribute Browser (2003)** (Marsden and Cairns, 2003). **TagsFS (2006)** also introduced file attributes and fluid hierarchies, where on each navigation step only relevant folders are shown to the user (Bloehdorn et al., 2006). A simple solution for files marked as unimportant and not ready to be deleted was provided by **GrayArea (2009)** as an intermediate area in every folder of a file manager (Bergman et al., 2009).

Usage in practice
User defined attributes within a database-like file system were fully supported in the discontinued BeOS operating system (now known as Haiku OS). To some extent, Windows Explorer in Windows 7 lets users define attributes (author and title) for certain file types such as Office documents, and it allows tags to be added to images. Since Windows Vista, it supports saving search results as virtual folders. The same feature is available in OS X Finder file manager. Such folders change dynamically based on the search query. Windows explorer also supports automatic grouping for certain file types such as stacking photos by month or music files by authors using the additional metadata of these file types (Exif, id3 tag). Other specific applications such as photo browsers and media players allow special attributes and tagging (e.g. people, geolocation, etc.) However these work in supplement to file browsers and will not be discussed here. There exist applications for OS X that allow single or cross-tool tagging of files, email, bookmarks based on OpenMeta platform (e.g. Tags, Tagit). Tags are integral part of the Finder file manager In the upcoming OS X version 10.9 (similarly implemented as OpenMeta). Tagging possibilities exist also in Gnome Nautilus or Nemo and KDE Dolphin (through Nepomuk – see 6.3) file managers.

4.2. Spatial management and piling

Malone argued that piles require less cognitive effort to be managed compared to categorizing information and that spatial management better served memory for locating documents as do hierarchies (Malone, 1983). The idea was prototyped for PIM by Mander et. al in the **Pile metaphor (1992)** (Mander et al., 1992). **Dynapad (2004)** introduced non stacked piles of photos' and documents' thumbnails on a zoomable interface to achieve higher visual recognition and remindability, and to overcome spatial limits (Bauer et al., 2004). It supported implicitly organized piles and clumps which could be created by encircling a set of thumbnails with a mouse pointer, sub-piling and time visualisation. **Archy (2005)** also provided a zoomable desktop where information (thumbnails) could be zoomed in and directly edited (no need for applications) and was always in a persistent state (Wikipedia, 2005). **Bumptop (2006)** enhanced piles by moving them to a 3D environment and introduced physics to piles' interaction (Agarawala and Balakrishnan, 2006).

Usage in practice
Even early file managers allowed spatial management of document icons (or thumbnails) in folders. This feature still exists for example in OS X Finder, Gnome Nautilus and other spatial file managers, while it was

disabled in Windows Explorer on Windows 7 (where auto-arrangement can not be officially turned off). Windows Explorer in Windows 7 also tries to hide a hierarchy behind so called libraries which aggregate multiple storage locations. During navigation it does not automatically expand hierarchy on the left pane by default, which hides it from the users (showing only the content of one folder at a time). It also introduced (as we mentioned above) arrangement views which let users stack files such as photos or music based on metadata (month, author) which are shown as piles in a folder. However, the Desktop spatial area can be freely managed on all major operating systems and file thumbnails and previews are also present.

4.3. Searching and associative browsing

Piles and spatial management provide contextual clues for retrieving and reminding. Organizing information was as important as retrieval in the above presented tools. A tool focusing on retrieval only, by providing a compiled list of related documents to a currently used one was **Remembrance agent (1996)** (Rhodes and Starner, 1996). Several indexing techniques were later developed and tested in file search engines such as **GLIMPSE (1994)** that provided full-text search, approximate matching allowing misspelling and support for regular expressions (Manber and Wu, 1994). **Dynamic queries (1998)** were also implemented on a top of a file system providing easy real time manipulation of information visualisation with widgets (sliders, buttons, etc.) (Liao et al., 1998). When these were moved and clicked, appropriate information was shown on the screen, although in the file system, only size and time could be manipulated. **The Logic Information System FS (2002)** allowed querying file attributes using a boolean logic similar to some web search engines (Padioleau and Ridoux, 2002).

mspace (2003) (Schraefel et al., 2003) was a multi-faceted bidirectional search interface for semi-structured information. A more general system that supported navigation of semi-structured information is **Magnet (2003)** which integrated keyword search and tags navigation, provided suggestions of similar results, and kept track of a browsing history (Sinha and Karger, 2005). **Connections (2005)** added context to the file search by combining content and time proximity of usage to form a list of results (Soules and Ganger, 2005). **Provenance file search (2007)** on the other hand combined content and document provenance or relations through system calls such as copy and paste (Shah et al., 2007), building on ideas from provenance aware file system **PASS (2006)** (Muniswamy-Reddy et al., 2006) (now being extended to the cloud). Linkage between files is also provided by personal ontologies over a file system such as **OntoFM (2010)** (Rompa et al., 2011). Another type of associative search and browsing was provided by **InfoATV (2012)** that helped re-finding through re-creations of previous search sessions together with showing a network of related documents based on associating personal documents with online collections (Madlock-Brown et al., 2012).

Usage in practice
Cross-tool indexing (not limited to file system only) and full text search are a standard in current operating systems, but often search engines are used only as a last resort when users forget where they had placed the information items (Bergman et al., 2008). However there is little or no support for semantics in file managers (one exception is KDE Dolphin with a semantic desktop which is presented in subsection 6.3), and additional context clues (more tools addressing this issue are presented in subsection 6.2). Although provenance (document evolution) was used here only for file search, present operating systems offer a possibility to navigate (or restore) to old file versions (more on provenance navigation in section 6.1). Note that versioning is just a subset of provenance as it does not include links to other sources. Windows supports this feature in a file manager though right mouse click and selecting *restoring previous version*, while OS X (since version 10.7 Lion) supports this in each application separately which is run automatically or manually from the application titlebar.

4.4. Time dimension

As we have seen (section 4.2), Dynapad provided a time scale view over information items. Time based ordering was used already in the **Perspective Wall (1991)** (Card et al., 1991) where search results were placed on a fish-eyed grid ordered by time and file types (Mackinlay et al., 1991) (a part of The information visualizer presented in subsection 4.6). **MEMOIRS (1992)** also displayed information (documents and diary entries) as events on a timeline interface supporting (1) episodic retrieval by exploiting autobiographical memory, and (2) retrieval by recognition (based on visual appearance) by allowing items to be colour coded (Lansdale and Edmonds, 1992).

TimeScape/TimeMachine (1999) exploited the time dimension as well in a different way. This was a time based desktop manager which preserved its past states, functioning as an automatic archiving system. It also allowed creation of future states that served as reminders (Rekimoto, 1999). Another time based application **Time2hide (2008)** (Lepouras et al., 2008) kept the computer desktop uncluttered by hiding old and unused icons (fading effect) using general and per icon hiding settings. The time dimension in unification prototypes (such as Lifestreams) is presented later in section 6.

Usage in practice
In all major file managers, a list of files can be ordered by time (similar to email or web browsing history where default organization is by time), size and type on a folder basis. Besides a list view, major file manages provide a thumbnail view as well. However, other visualisations and browsing possibilities are not present. One major exception is the Nemo Documents file browser, which shows files in a calendar interface based on creation, access and modification time. It also supports faceted search to address clutter on a week, month and year view (http://www.nemo-docs.com/).

4.5. Agents and automatic classification

Automation for filing information was used in a data mining prototype **Athena (2000)**. It could discover topics and organize files by them, reorganize documents by giving it some as example, route documents based on current folders' content and find misplaced documents (Agrawal et al., 2000). Interestingly there are no other prototypes to our knowledge that supported automated filing in file hierarchies on behalf of users. Perhaps this is because the file hierarchy grows incrementally and its growth is under users' control, which is not the case with email.

4.6. Different visualisations of the file storage

Several previously mentioned prototypes tried to change the visual perspective and management practices used in hierarchy file system. These were piles on a spatial surface (Pile metaphor, Dynapad, Archy, BumpTop) and time based grid view (Perspective Wall, MEMOIRS).

Several visualisation techniques were used to spot patterns such as size of files/folders or hierarchy depth. **Information cube (1993)** presented a hierarchy in a transparent nested box metaphor where each box represented a (sub-)folder (Rekimoto and Green, 1993). **Information slices (1998)** visualised hierarchies using cascading, semi-circular discs where clicking on a part of a disk revealed the corresponding part in a new zoomed in disk (Andrews and Heidegger, 1998). Another radial visualisation is **SunBurst (2000)** which radially lays items in a hierarchy, with the top of the hierarchy at the centre and deeper levels farther away (Stasko et al., 2000).

Treemap (1991) uses the size of files and visualises them as nested rectangles (based on folders) where the biggest file (folder) takes the biggest space on screen (Johnson and Shneiderman, 1991). A version of tree maps called Cushion Treemaps (1999) tries to present every rectangle as close to a square as possible (Van Wijk and Van de Wetering, 1999). A time based treemap called **Treemaps in time (2010)** visualised changes in the file hierarchy through time (Xu et al., 2010). Because treemaps are not showing the tree structure explicitly (e.g. non-leaf nodes cannot be pointed at) other variants were developed such as a 3D version **StepTree (2004)** (Bladh et al., 2004) and **Beamtrees (2002)** (Van Ham and van Wijk, 2003).

A common way to visualise large hierarchies is the use of a tree (graph) metaphor. **The information visualizer (1991)** was a 3D version of the Rooms virtual desktop (Henderson and Card, 1986) that also (besides the Perspective wall presented in 4.4) allowed users to visualise their information in a 3D hierarchical visualisation Cone Tree with a zooming interface (Card et al., 1991). Several other tree based visualisation techniques (whether used for PIM or not) to visualise large hierarchies are available such as a classic radial tree, **Fractal trees (1993)** (Koike and Yoshihara, 1993), **Fviz (1995)** – an improved cone tree or a balloon view of a radial tree (Carriere and Kazman, 1995), **Hyperbolic browser (1995)** (later Star Tree Studio) – a hyperbolic tree with a fisheye zooming capability (Lamping et al., 1995), **RF-Cone tree (1997)** – a 3D version of a tree similar to Fviz (Teraoka and Maruyama, 1997), **Cat-a-Cone tree (1997)** – a cone tree with additional tree information (Hearst and Karadi, 1997), **Disc Trees (1998)** – another 3D variant of a cone like tree (Jeong and Pang, 1998), **Circular trees (2006)** – similar to radial or hyperbolic tree (Ciccarelli et al., 2006), **Hierarchical Edge Bundles (2006)** (Holten, 2006), and **Bubble tree (2006)** (Grivet et al., 2006). There are also more abstract hierarchy visualisations such as **Botanical Trees Tree Viewer (2001)** (Kleiberg et al., 2001; van Wijk et al., 2003) and **PhylloTrees (2006)** (Neumann et al., 2006).

Another two visualisations are Content Map and ContactMap. The **Content Map (2009)** prototype provided an overview of a "file hierarchy" in a mind map like view (Espenkrona and Svensson, 2009). Authors of the social desktop ContactMap (see section 3.4) had planed to associate files with contacts, so files could be navigated to through a social network. Such navigation would probably be good for only a small number of files that would be connected with people in a contact list.

Usage in practice
Several of these visualisations were not meant explicitly for file hierarchies. Compared to email, the visualisation of the file hierarchy suffers from the lack of attributes on which it could be visualised; files do not have consistent (and sometimes meaningful) user friendly attributes (such as email – sender, receiver(s), time, subject, body, size, attachment). Some file types such as photos and

music provide more opportunities to visualisation with attributes as creation time, GPS location, author, genre, etc. For this reason, several tools specialized for managing only photos, music or videos emerged, but are beyond the scope of this paper.

There are some 3D file managers available. Probably the best known 3D file managers are SGI File System Navigator (made popular by the film "Jurassic Park") and **3DOSX (2002)** (Chin, 2002). However, there is no proof that a 3rd dimension brings any advantage over present 2D hierarchies. It could even be hypothesised that a 3D interface contributes to slower retrieval times if considering the research of spatial interfaces (Cockburn and McKenzie, 2001a; Sebrechts et al., 1999). Several treemap and sunburst (also multi-level pie or ring(s) chart) visualisations projects are available for all major operating systems as well (e.g. Gnome's Boabab, KDirStat, WinDirStat, Disk Inventory X, GrandPerspective, DaisyDisk etc.)

5. WEB BOOKMARKS AND HISTORY

Most present day web browsers have a web bookmarks manager and web history embedded. Regardless of this, people employ several other bookmarking techniques such as sending URLs by emails to oneself, copying them to personal web pages or documents, printing out web pages, saving them in a file hierarchy, etc (Jones et al., 2002). To improve the browsing and retrieving experience several research prototypes have been developed. This section first describes bookmarking prototypes, then history prototypes and finally prototypes that integrated both.

5.1. Web bookmarks

Some prototypes tried to impose metaphors from the physical world. Such examples are **WebBook and WebForager (1996)**. The first is a 3D book metaphor where pages of a book present web pages, while the second tool provides a 3D room-like environment to store and manage WebBooks (Card et al., 1996). **Data Mountain (1998)** used a 3D inclined plane to spatially manage thumbnails of bookmarks that assisted retrieval with a spatial position and visual clues of thumbnails (Robertson et al., 1998). The problem was again limited space and scalability. A 3D spatial information space such as that used in Data Mountain lead to slower retrieval times compared to a 2D system (Cockburn and McKenzie, 2001a; Sebrechts et al., 1999).

Web spy (2001) augmented bookmarks with alerts about changes on bookmarked pages (Sorensen et al., 2001). **TopicShop (2001)** helped users organize (clumsy and neat) piles and tag thumbnails on a spatial area, and evaluate them by providing web sites profiles (links to other sites in the collection, number of pages, images, multimedia, etc.) (Amento et al., 2000). **Tags (2003)** were also used in a multiple classification interface for managing bookmarks based on a flat set of attributes (Quan et al., 2003).

One of the tools supporting findings that users sometimes want to bookmark only a chunk (snip) of a web page and not all of it is **Snip!t (2003)** (`http://www.snipit.org/`). It lets users store, multi-categorize, share snips and suggests future actions if it recognises a type of data in a bookmarked text (e.g. a post code) (Catarci et al., 2007).

5.2. Web history

As with email, web history is automatically acquired and similarly attracted several researchers who tried to visualise it. **WebMap (1995)** visualised history as graphs where each page was represented by a small circle and types of links between pages indicated: (1) whether the destination document is located on a same server, (2) the navigation jumps to pages on different servers and (3) cross-jumps between nodes on the same server. The produced graphs could also be shared (Dömel, 1995). **MosaicG (1995)** depicted a graphical and zoomable left-to-right hierarchical history of web pages' thumbnails visited during a browsing session (Ayers and Stasko, 1995). **Padprints (1998)** similarly introduced a zoomable tree view of web browsing histories with thumbnails as nodes (Hightower et al., 1998). **WebPath (1998)** employed a 3D representations of web navigation, where pages were visualised as cubes (with thumbnail, title and other information on its surface) and connected with coloured arrows distinguishing domains (Frecon and Smith, 1998).

Domain Tree browser (2000) visually presented trees of browsing history thumbnails adjacent to the currently viewed page, organized by domains (Gandhi et al., 2000). Very similar was the **Global Tree browser (2000)** and both proved to be easier for retrieving history items (compared to standard list methods) with the use of visual aids (Killam, 2001). **Footprints (1999)** made use of real-world metaphors such as maps (a graph of web sites), paths (a tree view history), signposts (popularity of documents by colour), etc., to expose history of use by a group of users (Wexelblat and Maes, 1999). **Browsing icons (2001)** displayed animated graphs of users' browsing paths similar to Footprints' maps. Every web session built an individual graph which could be organized in a hierarchy of user-defined reusable and resumable tasks (Mayer and Bederson, 2003).

History-Centric Browsing (2006) associated pages from web history to the currently displayed page through three types of relevancies: temporal sequence, URL/location-based proximity, and content similarity. It could show these related pages during the browsing session through 3 different graphical visualisations: a statistical summary of a web history, thumbnails of relevant pages on top of the current web page, and a temporal exploration of history based on relevance to the currently visited page (Shirai et al., 2006). **Contextual Web History (CWH) (2009)** integrated several of the above ideas, introduced search and also inserted thumbnails of relevant web pages from the history on a Google's web site based on a search query (Won et al., 2009). Another web tool provided remindability of already visited web pages with more prominent web page thumbnails called **Visual snippets (2009)** that were composed of a web site logo, title and image (Teevan et al., 2009). **Trails (2011)** visualises web history on a spatial surface in different views taking advantage of tagging, tag clouds, shapes and colours of each item (Yu and Ingalls, 2011).

5.3. Integration of web bookmarks and web history

BookMap (2000) was a prototype that integrated bookmarks, web history (last few pages), most visited pages and pages marked to be read later in a 2D manageable graph (Hascoët, 2000). **WebView (1999)** allowed users to browse their web browsing histories in several modes: (1) a tree view based on nesting relationships between the storage location of the pages, (2) a tree view based on cross site navigational links and (3) temporal view. It also allowed visually annotating thumbnails with dogears (bookmarking) based on visiting frequency (Cockburn et al., 1999). **Integrated System (2001)** integrated the browsers back button, a browsing history, and bookmarks in one interface. A click on a back button showed recently visited web pages as thumbnails, which could be dog-eared (bookmarked) and filtered with dynamic queries (Kaasten and Greenberg, 2001).

Bookmark Pruner (2001) let users score bookmarks on an importance scale and used the web history (amount of time spent at each web page, frequency of revisits, and length of period between revisits) to automatically order and gradually change the list (Sorensen et al., 2001). Similarly **History Harvester (2001)** created a list of bookmarks from the whole browsing history as input and produced a list of web bookmarks based on user's predefined areas of interest (keywords) (Sorensen et al., 2001).

3D bookmark (2004) was a bookmark system allowing users to browse bookmarks and history (separately but in a consistent interface) in three listing modes: a book mode, a circular mode, and a cube mode. All three modes previewed previous and following pages, displaying several pages simultaneously (Yamaguchi et al., 2004).

Usage in practice

Present web browsers made bookmarking easier with one mouse click, better search tools over bookmarks and synchronizing bookmarks (whole profiles) between different devices (web browsers). Some browsers also allow tagging in conjunction with a hierarchy (e.g. Firefox). Tagging is also used in some standalone applications such as miTaggedMarks and on many social bookmarking sites such as Delicious which provide an online personal and collaborative classification and taxonomy of web bookmarks.

Web browsers mostly show a timely ordered list of previously visited web pages. They also gained a search feature and real time results when typing in browsers' address bar. Although a web browsing history lists lack organization, they still allow some maintenance (e.g. deleting items from a history list or private mode which disables logging browsing history). Other than this web history was (to our knowledge) not subject to any other visualisation even though studies suggest that users revisit anywhere between 60% to 80% of web pages (Tauscher and Greenberg, 1997; Cockburn and McKenzie, 2001b). It is also surprising that users use web histories in only a small percentage (in some studies even less than 1%) of revisits (Weinreich et al., 2006; Jones et al., 2001, 2002; Byrne et al., 1999; Aula et al., 2005). There are also some commercial products such as Browseback (`http://smilesoftware.com/BrowseBack/`), that let users visually scan thumbnails of web pages, or MindRetrieve (`http://www.mindretrieve.net/`) which allows users to search their web history, tag bookmarks and even files (although the development seems to be stalled).

The new Firefox feature called Tab Groups (previously known as Panorama or Tab Candy) brings a new visual way to manage web sites in groups on a spatial surface. Users can then only open tabs of web pages from a particular group. Future versions promise to support collaborative web browsing. Tab Groups is a kind of mashup between bookmarking and (collaborative) browsing (or performing tasks), using several ideas from many of the prototypes discussed in this paper (virtual desktops, zooming interface, piling, etc.).

6. INFORMATION INTEGRATION

Information fragmentation is considered one of the main obstacles in cross-tool and cross-device information management (Boardman and Sasse, 2004; Ravasio et al., 2004) and has since been a focus of the PIM research community. Some of the prototypes already described provided some level of integration. These include Raton Laveur (email, calendar, to-do list and notes – see 3.3) and TaskMaster (emails, URL's and attached files –

see 3.3). But their focus was on one information type within a standard interface while providing support for managing non native information types as well. Contact Map's idea was to integrate contacts with several types of communication (email, instant messaging, telephone) and files; although only email was implemented (see 3.4). On the contrary, this section provides an overview of information integration by other means. First we presented applications that provide integration through searching (6.1), followed by applications that integrate through context, time and tags (6.2), semantic desktops (6.3) and task support (6.4)

6.1. Search and associative browsing

Probably the best known and used cross-tool PIM applications are desktop search engines. One of the first such prototypes was **Stuff I've seen (2003)** (SIS) which provided a keyword and faceted search over several kinds of information types such as files, email, calendar entries, contacts and visited web pages. It also displayed rich contextual clues with returned results such as parts of the text and attributes (Dumais et al., 2003). It was later upgraded by **Milestones in Time (2003)** which showed results on a timeline together with public (e.g. holidays) and personal milestones (from calendar, photos) to support episodic memory (Ringel et al., 2003). SIS also lead to a new desktop search engine **Phlat (2006)** which merged search and browsing via various associative and contextual clues (people, tags, time, hierarchy path) (Cutrell et al., 2006).

Semex (SEMantic EXplorer) (2005) enabled browsing personal information by automatically created semantically meaningful associations (Dong, 2005). Retrieving information by association has lead to development of **Quill (2008)** (Gonçalves and Jorge, 2008) and **Feldspar (2008)** (Chau et al., 2008). Quill is a so called narrative-based interface where the user fills in predefined sentences (tells a story about a document) and the system returns a list of thumbnails of related documents. Feldspar is very similar but lets users interactively link predefined sets of information types presented in columns with associations. Similar associative browsing in a graphs is provided by Project Infospace (see 6.3), **Provenance graphs (2010)** (based on the TaskTracer - see 6.4) (Jensen et al., 2010) and **Versionset (2011)**, a copy-aware computer ecosystem (Karlson et al., 2011). These prototypes showed information evolutions through operating system calls in a graph like interface. Another prototype called **VisMe (2010)** allowed users to visually explore a map of personal information via relevant concepts in interconnected timelines (Gomes et al., 2010).

The use of associative memory search was adopted in the **xSearcher (2009)** which created associations (semantic links) between information items from explicit and implicit user activities and visualised them in a graph (Chen et al., 2009). **Poyozo (2010)** allows time proximity (as Milestones) search in a calendar interface based on tags or keywords which are automatically aggregated from personal information of one's online activities (such as email, browsing history, social networks, blogs, etc.) (Moore et al., 2010). Time proximity search is also used in the **YouPivot (2011)** search engine that visualises the results in a so called modelled activity and allows annotation of important moments in history (milestones) (Hailpern et al., 2011).

Usage in practice
Although powerful desktop search engines are a part of our computer lives, they are rarely used to access information (Bergman et al., 2008). These search engines are good at retrieving a particular item given a correct set of search keywords. However, retrieving all information items related to e.g. a particular project with one search query is almost impossible or requires a high cognitive load. An upgrade to today's desktop search engines is Beagle++ that ranks search results based on the number of semantic links (a part of the Nepomuk semantic desktop - see 6.3). Associative retrieval through stories can take a lot of time with present interfaces and it probably has a prominent future integrated with better voice recognition. Provenance is still not supported in the mainstream operating systems, however a lot of research is going on in this direction at the moment. The Poyozo visualisation is available as a web browser extension while YouPivot is just being offered as a product to the public.

6.2. Context by time proximity, tags and dynamic structures

Time proximity was a core in several above mentioned search and visualisation tools but without organisation possibilities. Time organization and context provision was used in **Lifestreams (1996)**. It showed several types of information items (files, email and calendar inputs) in a timely ordered document stream which could be managed, organized and summarized by applying different filters to it. Old documents moved towards the tail of the stream (simple archiving) and documents could be placed in the future as reminders (Freeman and Gelernter, 1996). **Webtop (2002)** showed a list of documents related by various types of relationships (the same folder, bookmarks, content, etc.) to a currently opened document (Wolber et al., 2002). **Implicit query (2004)** also generated a list of related information items to the users' current activities by context-sensitive search (Dumais et al., 2004) while **Ivan (2008)** created a list based on time proximity (Pedersen and McDonald, 2008).

Presto (1999) allowed spatial management of information and their (automatically or user defined) attributes (tags) in piles. Dynamic groups could be formed by search queries or manually by users and could be further refined by dragging additional attributes onto them which would change their content accordingly (Dourish et al., 1999). **KnowledgeSpace (1999)** extended a file limited DomainView (see 4.1) by allowing users to set tags on other information types as well (files, emails, contacts, web pages, etc.). Tagging multimedia (video and images) on a hard drive and web was provided by **M-Refind (2012)** that allowed users to use their regular PIM tools for tagging with the addition to the context-aware retrieval system (Zhao et al., 2012). A step forward in dynamic documents' structures and context provision (though various means) are desktop applications such as Haystack, Gnowsis and its successor NEPOMUK which will be presented later (see 6.3).

Usage in practice
Unification through time, tags or context is not present in mainstream information managers. Tagging is supported in several applications separately, which use different approaches as there is no standard to follow. Lifestreams was released as a commercial application called Scopeware, but it never became widely used as a desktop application. One drawback of Lifestreams is that it replaced one superordinate aspect of the document (its location in the hierarchy) with another (its location in the timeline). However, "lifestreams" are used by several other tools such as social networking web services and blogs. Social networks aggregators also unify information from different social networks and present them in a time ordered stream of posts. Similar aggregation was achieved in Vodafone 360 (discontinued) for all communication streams of contacts from a unified contact list (email, social networks, phone).

6.3. Semantic desktop

Project Infospace (2003) replaced a desktop metaphor and implemented several ideas already mentioned (multiple classification in e.g. various contexts, document evolution or provenance (see 6.1)) with bi-directional links between documents and groups of documents to manage semantic associations and access related information (Ravasio et al., 2003). **Haystack (1999-2005)** also provided its own interface and a unified namespace for one's data environment. Information items (or even parts of them) could be grouped, annotated, and linked in any possible way, giving users total control over the organization (Karger et al., 2005).

IRIS (2005) on the other hand used existing tools and integrated them in a coherent interface, allowing users to annotate, classify and display related information in a familiar environment (Cheyer et al., 2005). **DeepaMehta (2005)** linked information from separate desktop applications into a graph-based user interface (Topic Map). It let users organise, describe and relate information objects, and navigate graphs (Richter et al., 2005). Prototype **gnowsis (2003-2007)** also complemented the current desktop with semantic web features (Sauermann et al., 2006b). In addition to providing an interface for managing personal information, it also provided interfaces for other applications to access this information. The personal ontology used in gnowsis was also used in its successor **NEPOMUK (2007-2009)** (Groza et al., 2007). This semantic desktop system introduced collaborative work and communications by exploiting existing applications.

Usage in practice
To foster standardization and interoperability between different implementations, the community around the NEPOMUK project initiated the OSCA Foundation whose standards are now used by KDE, gnome, freedesktop and former Nokia's Maemo. NEPOMUK is being integrated in the KDE - K desktop environment (which is available for all major operating systems including Linux, Windows and OS X). Although the development is slow, it is already possible to tag, rate, and comment files in a KDE file manager (Dolphin) and search this information with search engines (Strigi and Gnome's Meta Tracker). The future versions plan to extend semantic desktop features to all KDE applications. The DeepaMehta project is also still under development (`http://www.deepamehta.de/`) while the development of Open IRIS prototype has ceased (`http://www.openiris.org/`). A KDE project worth mentioning here is Akonadi. Although not related to semantic desktops, it provides a unifying storage and retrieval services of data and metadata to every PIM application that traditionally have different data storage and handling methods.

6.4. Project and task support

High-level task (also referred to as projects, activities) management has long been a focus of researchers who produced several different approaches to recreate the task context (see also 3.3). One of the first prototypes that addressed this was the **Personal Role Manager (1994)** (Shneiderman and Plaisant, 1994). It allowed users to manage multiple information types under so called roles (and associated tasks), represented as a spatially arranged hierarchy of windows. **WorkspaceMirror (2003)** helped users maintain similar hierarchies across applications with mirroring folders between them. Although the prototype was not meant to be specifically used for tasks, researchers found out that mirrored folders tend to be task (project) related (Boardman et al., 2003).

UMEA (2003) allows users to define and select tasks. All information items used were then automatically associated to the manually selected task and ordered based on the time spent working on each information item (Kaptelinin, 2003). A similar environment was provided by **Activity Explorer (2004)**, combining personal and collaborative task management (Muller et al., 2004). **Sphere Juggler (2004)** allowed manual assignment of information items to tasks in a hierarchical list of task related information divided by type, and automatically monitored task switches (Morteo et al., 2004). In a similar fashion **TaskTracer (2005)** monitored users' activities and documents usage and with user assistance associated these activities to tasks (Dragunov et al., 2005). The prototype gained automatic task prediction with a TaskPredictor (Shen et al., 2006) that automatically detected low-level task shifts in support of larger tasks. Automatic spatial categorization of task (project) related information in circles was provided by **CAAD (2007)**, which in addition allowed users to manually move information between tasks (Rattenbury and Canny, 2007).

TaskVista (2006) was a to-do list which could be created by typing in to-dos or by dragging-and-dropping emails or files to it. It was a part of a bigger application called ACTA (Activity-Centered Task Assistant) which created a task centred environment with separate folders for documents, attendees, correspondence, contacts etc. (Bellotti and Thornton, 2006). The use of personal ontologies was (besides in semantic desktops) also exploited in **DELOS Task (2007)** – a task-centred information manager – to automate the user's most frequent activities and compute task inferences (Catarci et al., 2007). **Giornata (2009)** provided task specific virtual desktops where each desktop contained task specific (shortcuts to) files, people (and groups), tags and applications (Voida and Mynatt, 2009)

Personal Project Planner (2005) or PPP (previously referred to as Universal Labeller (Jones et al., 2005) and later as PlanZ (Jones et al., 2010)) let users write a project plan in a word processor and automatically turned headings into a file hierarchy in the background. Text from web pages and emails could be dragged-and-dropped in a document, emails could be written, and links to original sources were automatically created making it possible to return to original content and context of information. The granularity of information (using smaller chunks instead of whole information items) was also addressed in Snip!t (see 5.1) and Haystack (see 6.3).

Usage in practice

Ideas from UMEA and TaskTracer were later used in several time tracking applications (e.g. Chrometa). Some task support was included in Windows XP Explorer via a task pane on the left side of the window containing a list of common actions and destinations that were relevant to the current directory or file(s) selected. While it did not integrate information, it provided some online services in a file manager. This was removed in Windows 7. PPP is available for Windows users as PlanZ. There are plentiful project management applications (e.g. see `http://en.wikipedia.org/wiki/List_of_project_management_software`) which focus more on collaborative project aspects and not PIM.

7. DISCUSSION

We categorized 146 PIM research prototypes on personal computers. The aim of this categorisation was to see whether the two common beliefs are true: (1) not many of PIM prototypes have been extensively evaluated (Bergman et al., 2004; Boardman, 2004) and (2) a lot of radical design is not making it into the real world (Bergman et al., 2004; Whittaker et al., 2000). The second belief has been partially challenged by a short discussion after each subsection where we looked at ideas and technologies that have been (to various extents successfully) transferred, replicated from the research prototypes to or used by publicly available applications. We will return to this issue after trying to gauge how much evaluation of PIM tools has been done.

All discussed prototypes are listed in Figure 1. They are presented in a chronological order based on the first appearance in the academic literature divided by information type: email, files (split in two columns for equally spacing the table), bookmarks and history, and unification. For every prototype we tried to find a reported user study in the original paper or other papers that cite the original one. The number of participants, a short description of each study and the number of citations for each paper (from Google Scholar on the 20.9.2012) are listed by every prototype. The last column presents if authors of the paper are from academia or industry; "A" means that authors are from academia only,"I" from industry only, "AI" that the first author is from academia but some are from the industry, and "IA" the other way around. In brackets are companies at which authors from the industry worked at the time of publication and that appeared in more than one paper: X – Xerox or just Palo Alto, A – Apple, I – Intel, M – Microsoft, L – Lotus, H – HP, G – Google and O other.

For several prototypes we could not find a reported user study. Some prototypes were studied just from a technological point of view (e.g. comparing algorithms), while some had several studies under different names (e.g. Mailcat (Segal and Kephart, 1999) as Swiftfile (Segal and

Kephart, 2000), PPP (Jones et al., 2008) as PlanZ (Jones et al., 2010) and Extreme Blue as reMail and Dremail (Kerr and Wilcox, 2004)). For some prototypes studies of different features were published separately (e.g. Haystack (Karger et al., 2005; Quan et al., 2003), TaskTracer and Provenance (Dragunov et al., 2005; Jensen et al., 2010), and Thread Arcs, reMail, Dremail and Extreme Blue (Kerr, 2003; Kerr and Wilcox, 2004)). The number of participants by each prototype is the number reported in the study of a referenced paper. If more that one study was reported in the same paper (MailStacker, Padprints and Quill), the numbers were summed together. The number of participants was added also for The Bluemail for which we found two studies in two different papers (Tang et al., 2008; Whittaker et al., 2011). For the gnowsis prototype the same study was reported twice (Sauermann et al., 2006a,b), hence the two numbers of citations count. Two citation numbers are also listed by PPP/PlanZ as the tool is mostly know by its former name (based on the number of citations), while the evaluation study was reported under the new name.

7.1. "Prototypes are not extensively evaluated": common belief 1

The distribution of user studies (65 reported) can be seen in Figure 2. User studies of PIM prototypes were rare before 1995. Before the year 2000 nearly all user studies were done on predefined tasks in a lab. Some prototypes were evaluated on user datasets or performance benchmarks. During this period only separated studies of usage of email, files and web bookmarks were present. The interest in user studies of unification prototypes started only a decade ago. After 2000, more than a half of the prototypes (56 out of 104) were evaluated with users to various extents – from short laboratory studies to longitudinal evaluations of usage. This is 19 out of 28 email prototypes, 20 out of 31 unification prototypes, 12 out of 29 file prototypes and 9 out of 16 web bookmarks/history prototypes. Of all these studies, 39 were done in the participants' own environment compared to 18 lab studies. This trend of studying PIM in the real world is following the suggestions for the need of such studies from the emerging PIM community at the beginning of the century (Bergman et al., 2004).

A few patterns emerged in the graphs on Figure 2. On the email only graph (dark gray dots), the number of participants in studies grew each year. Email has become ubiquitous and it is almost impossible to imagine knowledge work without it. On the other hand the interest in the studies of the web bookmarks/history prototypes (empty circles) almost ceased after 2003. Newer studies show that the page re-visitation through bookmarks is very low (at about 10%) and it happens through new interfaces (e.g. bookmark toolbar) besides the bookmark hierarchy (Weinreich et al., 2008). The web has also become more service oriented and filled with dynamic content compared to a decade ago which is reflected also in the shift of research interest to sharing web content on social networks and social bookmarking (a search for "social bookmarking" on Google scholar returns 57,000 results from 2003 to February 2013, compared to a few hundred when searching for combinations of words "web" or "internet" and "bookmarks", "shortcuts" or "favourites").

Based on the numbers alone, less than half (65) of the described prototypes (146) reported a user study and for the majority of these it seems that the development stopped after the first study and first published paper (we already mentioned 5 prototypes that we know were tested in more than one user study in the introduction of this section). Although user studies (especially those done in the users' own space) are more common in the last decade, there are still several prototypes that have not been evaluated even in recent years (see the table in Figure 1). After 2005 only 30 prototypes have been tested with users while 25 have been not.

In papers that reported a user evaluation the number of participants per study is not growing in time (see the top graph in Figure 2). Rather it seems very stable. We only came across six very large studies: GrayArea 96 participants (files), Phlat 225 and SIS 234 (search engines in unification), Bluemail 345 & 15 in a preliminary study (email), and Visual snippets 276 and PadPrints 73 participants (web prototypes). The four biggest studies were done in the industry (Microsoft and IBM). If counting in these six large studies, the median for the number of participants is 12 ($\sigma = 64.6$). If excluding large studies, the median drops to 10 ($\sigma = 9.7$). Overall, the number of participants per study is not high.

The number of papers divided by the number of participants and whether authors were from academia or industry can be see in the table on Figure 3. As with papers with authors from academia only, the majority of papers written by authors from industry (or industry and academia combined) reported on a study with less then 10 participants. We cannot claim that authors from industry reported on larger user studies as a few papers from large companies with research departments have either not reported any study or reported on a study with a small number of participants. However, bigger companies have easier access to a bigger user set as is clear from these 4 largest studies. The number of participants also depended on whether a PIM prototype collected only log data or it involved a qualitative observation. The majority of the prototypes were evaluated from the qualitative perspective and there is clearly a need for quantitative studies (all six large studies were quantitative in nature)

Email		#	Study nature	Citations	A or I
1987	Information lens	na	37 users in preliminary interviews on PIM	746	AI
1992	Tapestry	na	no evaluation reported or found	2259	I (X)
1993	Mona	na	no evaluation reported or found	13	A
1994	MaxiM	na	no evaluation reported or found	3147	A
1997	TimeStore	5	3 weeks	39	A
1998	Apple data detectors	na	internal research	174	I (A)
1998	CAFE	na	12 users in preliminary study	53	A
1998	Re:Agent	na	performed on a dataset	201	A
1999	Mailcat	6	used inboxes of 6 users	212	I (I)
2000	Raton Laveur	na	no evaluation (design by interviews)	123	I (X)
2000	SwiftFile	5	used inboxes of 5 users	90	I (I)
2001	Thread-based email client	6	performed on a dataset of 200 emails	143	I (M)
2002	Bifrost inbox organizer	10	1 month	104	I (L)
2002	Sudarsky's prototype	na	no evaluation reported or found	23	I (S)
2002	TaskView	21	2 sessions in a experimental setup	60	A
2002	Zest	na	no evaluation reported or found	20	A
2003	Personal map	15	log over a week	31	I (M)
2003	Taskmaster	9	a few days of usage and interview	338	I (X)
2004	Contact map	15	15 users few tasks of which 6 users 3 days	69	AI (O)
2004	Email Nodes	na	no evaluation reported or found	4	A
2004	EzMail	12	2 groups, between subjects, predefined tasks	17	A
2004	FaMailar	na	no evaluation reported or found	10	A
2004	PostHistory	10	ethnographic study	92	AI (O)
2004	remail	na	internal studies	37	I (I)
2004	Social Nets Fragments	10	ethnographic study	92	AI (O)
2005	EmailViz rhythm ...	1	visualisation on one person's archive	36	A
2005	MailStacker	8	5 initial 30 min, 3 for major study of 1 week	1	A
2005	MailView	na	no evaluation reported or found	19	A
2005	Mountain	na	no evaluation reported or found	na	A
2005	Threadmap	na	no evaluation reported or found	1	A
2006	EmailViz treemap ...	8	1 hour interview, visualisation on users' data	33	AI (M)
2006	Themail	16	visualisation on users' archive, interviews	127	IA (I,H,X)
2007	Anymails	1	author	na	A
2007	Mail Garden	na	no evaluation reported or found	na	A
2008	The Bluemail	360	1st 15 users 1 hour observed, 2nd logs 345 user	17, 12	I (I)
2009	Magnetmail	30	case scenario task	2	A
2012	SmartOffice	7	interviews	na	I (O)

Unification		#	Study nature	Citations	A or I
1994	Personal role manager	na	no evaluation reported or found	40	A
1996	LifestreaM	na	no evaluation reported or found	307	A
1999	KnowledgeSpace	na	no evaluation reported or found	13	A
1999	Presto	na	no evaluation reported or found	260	I (X)
1999	Haystack	na	studies on different parts of the system	184	A
2002	WebTop	na	no evaluation reported or found	21	A
2003	Milestones in time	12	participants' own workplace	138	AI (M)
2003	Project InfoSpace	na	just low level prototype develped	6	A
2003	SIS	234	6 weeks log data	658	I (M)
2003	UMEA	8	2 to 6 weeks	212	A
2003	WorkspaceMirror	8	average 44 days	57	A
2004	Activity Explorer	33	100 days of usage	80	I (I)
2004	Implicit query	na	no evaluation reported or found	54	I (M)
2004	Sphere Juggler	4	short usage	15	A
2005	DeepaMehta	na	no evaluation reported or found	31	A
2005	IRIS	15	few weeks (members of research group)	74	I (SRI)
2005	Semex	na	compared algorithM	79	A
2005	TaskTracer	na	experiments of different parts of the system	191	A
2006	Gnowsis	8	different evaluations for each version	79, 20	A
2006	Phlat	225	8 months of logs	179	I (M)
2006	TaskVista	na	no evaluation reported or found	1	I (X)
2007	CAAD	10	a week of logs of 8 users, interviews	15	A
2007	NEPOMUK	na	several small studies of parts of the system	119	A
2007	DELOS Task	na	no evaluation reported or found	18	A
2008	Feldspar	8	experiment with redefined data	39	A
2008	Ivan	~5	couple of months	9	IA (G)
2008	PPP (PlanZ)	8	PlanZ in 2010, own projects, up to 12 days	47,18	A
2008	Quill	31	1st study 10 and 2nd study 21 users	20	A
2009	Giornata	5	average 54 days	10	A
2009	xSearcher	na	no evaluation reported or found	7	IA (O)
2010	Poyozo	na	no evaluation reported or found	1	A
2010	VisMe	20	17 tasks, a predefined set of documents	3	A
2010	Provenance graphs (TaskTracer)	17	2 months	15	A
2011	Versionset (copy-aware system)	16	12 weeks usage, interviews	1	AI (I)
2011	YouPivot	7	compared traditional and contextual search	8	AI (G)
2012	M-refind	na	no evaluation reported or found	0	A

Files		#	Study nature	Citations	A or I
1991	Perspective wall	na	no evaluation reported or found	880	I (X)
1991	Semantic File System	na	run over a file storage of multiple users	502	A
1991	The information visualizer	na	no evaluation reported or found	591	I (X)
1991	Treemap	na	no evaluation reported or found	1102	A
1992	MEMOIRS	16	lab evaluation	66	A
1992	Pile metaphor	10	one hour session of 5 predefined tasks	272	I (A)
1993	Fractal trees	na	no evaluation reported or found	99	A
1993	Information cube	na	no evaluation reported or found	149	A
1994	GLIMPSE	na	no evaluation reported or found	442	A
1995	Fviz	na	no evaluation reported or found	145	A
1995	Hyperbolic browser	na	no evaluation reported or found	1090	I (X)
1996	Remembrance Agent	na	no evaluation reported or found	548	A
1997	RF-Cone tree	na	no evaluation reported or found	20	IA (O)
1997	Cat- a-Cone tree	na	no evaluation reported or found	278	IA (X)
1998	Disc Tree	na	no evaluation reported or found	77	A
1998	Information slices	na	no evaluation reported or found	168	A
1998	Dynamic queries	18	90 minutes session, predefined tasks	6	A
1999	DomainView	na	no evaluation reported or found	13	A
1999	Timemachine	na	research group	252	I (S)
1999	Cushion Treemaps	na	no evaluation (design by interviews)	251	A
2000	Archy	na	no evaluation reported or found	na	A
2000	Athena	na	compared algorithM	109	I (I)
2000	SunBurst	32	16 tasks, between subjects, predefined set	147	A
2002	3DOSX	na	no evaluation reported or found	na	A
2002	Logic file system (LISFS)	na	benchmark experipents	72	A
2002	Beamtrees	na	no evaluation reported or found	15	A
2003	Botanical Trees	na	no evaluation reported or found	15	A
2003	Attribute Browser	na	no evaluation reported or found	39	A
2003	Magnet	18	predefined tasks on a set of data	54	A
2003	Mpace	24	half an hour completing predefined tasks	32	A
2004	Dynapad	2	12 hours (from thesis)	18	A
2004	StepTree	20	between groups study 2 navigations	53	A
2005	Connections	6	six months log of 3 to 5 queries per user	127	A
2005	Linking file system (LiFS)	na	no evaluation reported or found	48	A
2006	Bumptop	6	one hour session of 29 tasks	190	A
2006	TagsFS	na	no evaluation reported or found	47	A
2006	PASS	na	benchmarking overhead	221	A
2006	Circular trees	na	no evaluation reported or found	764	A
2006	Hierarchical Edge Bundles	na	no evaluation reported or found	370	A
2006	Bubble tree	na	no evaluation reported or found	39	A
2006	PhylloTrees	na	no evaluation reported or found	18	A
2007	Provenance file search	27	one month	36	AI (H)
2008	Time2hide	8	one month	1	A
2009	Content Map	na	no evaluation reported or found	1	A
2009	GrayArea	96	maintenance tasks	12	A
2010	Visualizing treemaps in time	na	performed on a dataset (shared drive)	121	AI (M)
2011	TagStore	18	setup tasks 2 groups	3	A
2011	OntoFM	na	no evaluation reported or found	na	A
2012	InfoATV	na	no evaluation reported or found	na	A

Web		#	Study nature	Citations	A or I
1995	MosaicG	na	no evaluation reported or found	157	A
1995	WebMap	na	no evaluation reported or found	148	A
1996	WebBook, WebForager	na	no evaluation reported or found	480	I (X)
1998	Data mountain	32	3 groups predefined tasks, compared	502	I (M)
1998	PadPrints	73	37 in the 1st & 36 in the 2nd experiment	206	A
1998	WebPath	na	no evaluation reported or found	45	IA (O)
1999	Footprints	40	20 each condition, available to the public	446	A
1999	WebView	20	for preliminary study	88	A
2000	Bookmap	na	no evaluation reported or found	13	A
2000	Domain Tree browser	25	4 users predefined task (21 w/ GlobalTree)	38	A
2000	Global tree browser	21	compared with Domain tree browser	7	A
2000	TopicShop	40	20 per condition (pilot study before)	33	IA (O)
2001	Bookmark Pruner	na	no evaluation reported or found	20	AI (O)
2001	Browsing icons	10	training and one revisit in a week time	22	A
2001	History Harvester	na	no evaluation reported or found	20	AI (O)
2001	Integrated system	na	no evaluation reported or found	92	A
2001	Web spy	na	no evaluation reported or found	20	A
2003	Snip!t	na	no evaluation reported or found	18	A
2003	Tags	21	experiment of a predefined data	56	A
2004	3D Bookmark	na	no evaluation reported or found	8	A
2006	History-Centric Browsing	na	no evaluation reported or found	10	IA (O)
2009	Contextual Web History	12	2 one hour sessions a week apart (pilot)	27	A
2009	Visual snippets	276	1st phase and 197 2nd phase in 2 days	50	I (M)
2011	Trails	5	compared predefined tasks	0	A

Figure 1. A table of all described prototypes by information type with a number of participants, a short study description, number of citations (from Google Shcolar on 20.9.2012), and if the authors of the paper are from academia or industry. "A" means that authors are from academia only, "I" from industry only, "AI" that the first author is from academia but some are from the industry, and "IA" the other way around. In brackets are companies that appeared in more than one paper: X – Xerox or Palo Alto, A – Apple, I – Intel, M – Microsoft, L – Lotus, H – HP, G – Google and O other.

as PIM research needs to move from the era of observation or exploratory studies only to the era of larger quantitative explorations as well.

For prototypes that reported a user study we used a linear model to look at the possible effect the number of participants, length of a study (in days) and data used

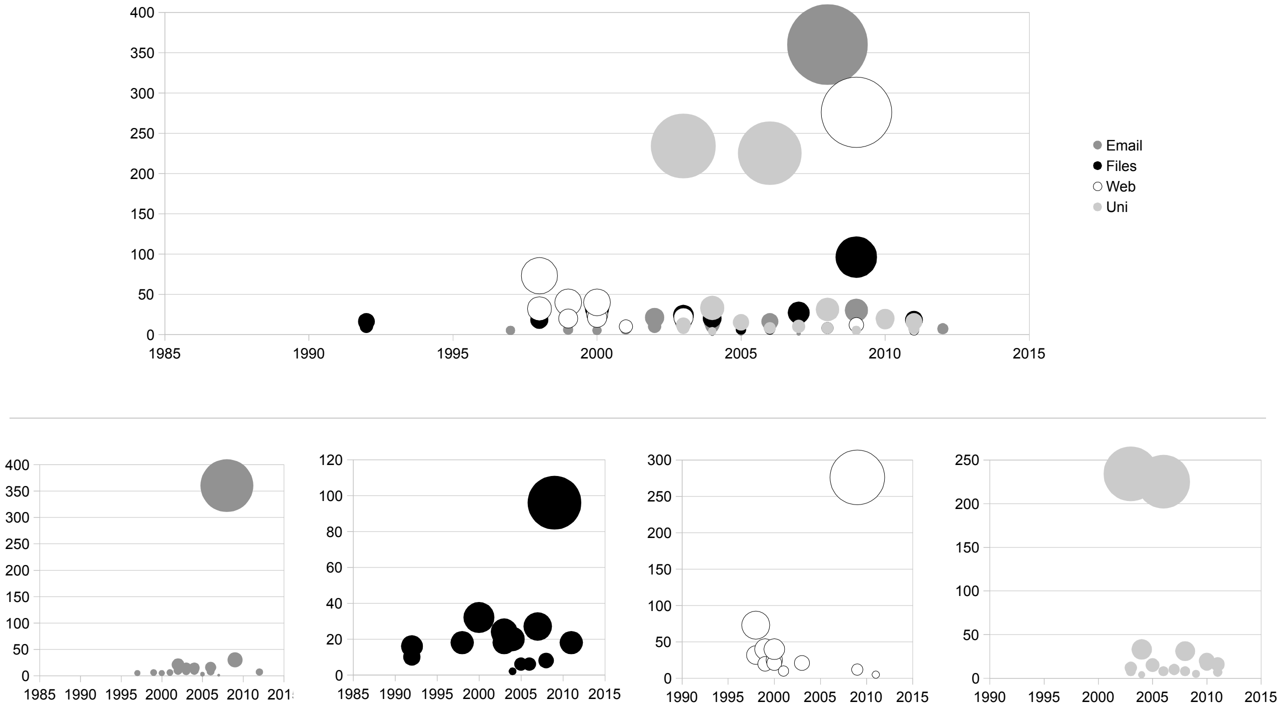

Figure 2. Number of participants per study by years. Size of each bubble is relative to sizes of other bubbles based on the number of participants in a study: bigger the bubble higher the number of participants. Top: all information types together. Bottom from left to right: email, files, web and unification prototypes.

Number of participants	na	0 < x <= 10	10 < x <= 20	20 < x <= 30	30 < x <= 40	40 < x	Sum
Papers with authors from academia only	58	18	11	5	3	2	97
Papers with authors from industry only	14	7	2	0	2	4	29
Papers with the first author from academia and some from industry	4	4	2	2	0	0	12
Papers with the first author from industry and some from academia	5	1	1	0	1	0	8
Sum	81	30	16	7	6	6	146

Figure 3. A table of papers divided by number of participants and whether authors were from academia or industry.

in a study (real versus lab) might have on the number of citations, which can be used as a measure of acceptance in the academia. Logarithmic values were used for citations, participants and days of the study. Looking separately on how each variable affects the number of citations, only the number of participants (with no outliers) had significant effect (Beta = 0.499, p = .0145), while the number of days and the type of a study (real, lab) were not statistically significant. However, there are so many counter-examples that its impossible to predict such relations. Some papers were highly cited with (to our knowledge) no reported user study and their ideas are implemented in everyday applications (e.g. email agents Information lens, Tapestry, Maxims). Some papers were highly cited, had been extensively studied and had made it into operating systems (e.g. Stuff I've Seen). However there are many more that were highly cited, but without an apparent full implementation in the real world (WebBook and WebForager, Data mountain, Pad prints, Web view, Lifestreams, Task master, Perspective wall, Semantic file system, Timemachine, Presto, UMEA) – although bits from many of these prototypes can be found in present PIM tools (e.g. thumbnails, tags). This shows that even if the number of citations is statistically correlated to the number of particiapnts in the study, it might be affected by many other factors: e.g. the time passed from the publication, acceptance of the paper by the community (e.g. conference awards), type of publication (scientific magazine, journal) and its impact factor, authors' visibility and popularity, ground-breaking ideas, results and even marketing behind some prototypes of authors from the industry).

The belief that *"not many PIM prototypes are extensively evaluated"* still holds. The average prototype has only one reported short-term study with about 10 participants limited to students, researchers and employees from companies and institutions where researchers worked. As noted before, it is hard to recruit people for PIM studies in their own environments because these involve potentially exposing private data (Kelley and Teevan, 2007). For long term studies (possibly involving many participants) the issue is even worse as such studies take time and personal information is a very valuable resource that needs to be managed in a very robust environment, which is often not possible to develop by researchers. Users are also known to resist novel applications if there is no clear or immediately visible benefit in using them, if these require a steep learning process or if they greatly change the established and trusted work practices. The smooth integration into current tools and practices is also often hard on the existing platforms (for example file managers, some web browsers or corporate email clients) as they are not very open to augmentation. Even if the belief of this section is true, we must take into account that many of the discussed prototypes were not meant to be used by users in the first place and that this claim needs to be looked at from a wider point of view as well. We will discuss this in the next section.

7.2. "Transferance of prototypes into the real world not high": common belief 2

If we look at individual prototypes, we can immediately come to the conclusion that the vast majority was never adopted in the real world. However, the transference of PIM research prototypes to the mainstream software seems to be more complex. There is a trend in science and technology indicating that inventions that attracted a critical mass of researchers creating and evaluating multiple versions of it, had greater chances of success (Whittaker et al., 2000). This can also be observed in PIM research if we look at some of the discussed technologies and solutions, rather than individual prototypes.

One example of successful technology transition is information indexing and desktop search engines (see 4.3 and 6.1) that evolved from public information retrieval solutions and attracted many researchers in the field of information retrieval from academia and industry (e.g. Microsoft). Cross tool desktop search engines are today a part of all major operating systems (in OS X from 2005 and in Windows from 2006). Email threads also received a considerable focus with one of the first prototypes developed in 1993 (Mona) followed by many others (see 3.2) developed and tested in academia and industry (e.g. IBM). Besides threading, email filtering, rules and agents for detecting spam are integrated into all major email clients. Tagging or multi-categorisation received significant attention as well, either evaluated separately (TagsFS, Tagstore) or as a part of a bigger system (Haystack, Presto, etc.) Tags are implemented in many PIM applications (e.g. Gmail, Firefox bookmarks), although only enhancing existing hierarchical organisations (even Gmail nested labels can be treated as folders), which is consistent with empirical studies that users are unwilling to use tags-only systems (Civan et al., 2008; Quan et al., 2003).

Alongside widely disseminated solutions, many small and helpful applications were developed based on the research ideas, although they never gained a broad user-base – e.g. file hierarchy visualisation software (see discussion in 4.6). The same is true for many discussed ideas that were implemented as extensions or add-ons for PIM software (e.g. Chrome, Thunderbird, Outlook, Gmail, Firefox, etc.) However, even in these cases the implemented ideas had a long evolution process (treemaps, sunburst, visualising email threads, social networks, etc.). Surprisingly rare are PIM research prototypes that try to directly evolve into commercial products. The ones that did, have either not succeeded (e.g. Lifestreams, Bumptop) or we do not know how widely they are used (e.g. Mailcat as an extension of Lotus Notes). Some prototypes were already a part of commercial applications when they were described in academic literature (e.g. Apple Data Detectors). Little is known about in-house user studies done in the companies that implemented these products.

Most of these successfully transferred technologies had either:

(i) a slow and smooth integration into present application environments without interfering with the current trustworthy user practices, which is very important for user adoption (Voit et al., 2009; Whittaker et al., 2000), and/or
(ii) were directly or indirectly supported and exposed by major software companies.

The majority of these technologies (desktop search engines, email threads and tagging, email agents, extensions of PIM tools) were just added to existing tools as augmentations and have not tried to replace them. Direct exposure happened in operating systems, widely spread PIM desktop software (Windows, OS X) or web platforms (e.g. Gmail) while indirect exposure took place in online software repositories (e.g. extensions) which was crucial for early adopters who spread the word of useful products further.

Besides (to various extent) successful adoptions, several technologies have not yet made it to the desktop even if they received significant attention by researchers. One of these is for example the semantic desktop (for which we could not find many user studies). Similar are provenance, context and associative information retrieval. These are fairly "new" technologies and we might have to wait longer for them to become a part of our daily lives. Task (or project) information management has also received significant interest from the research community. However, the proposed solutions to the problem differ significantly from each other and do not (clearly) present a continuing research approach (integrating information in email clients such as in Raton Laveur or Taskmaster, word processor of PlanZ, Personal Role manager for managing windows of roles, Workspace Mirror for consolidating hierarchies, etc.) There were also several prototypes that did not go through many iterations (mostly one), and even though supported by major companies, have also not made it in any form into commercial products (e.g. Web book, Web Forager, Data Mountain, Visual snippets, etc.)

Visualisation prototypes have also attracted attention in the research community and were welcomed in user studies, yet they are not available in mainstream desktop applications. Although visualisations are mostly not meant for keeping (acquiring) and organising information (see PIM sub-activities in (Boardman, 2004; Jones, 2007; Whittaker, 2011)) and thus cannot entirely replace current PIM tools (Kobsa, 2004), they can to some extent support maintenance and retrieval (e.g. visualising web browsing and consequently time management, spotting large files, folders with large numbers of files, hierarchy structure, social network structure, email conversations patterns). More importantly visualisations support "making sense of information" PIM subactivity (Jones, 2007) to reflect upon organisational structures, spotting events, changes, and oddities in information management. Users would certainly benefit from visualisations as an augmentation to current tools as they do not interfere with current practices.

An important question is *what was the goal or motivation behind PIM prototypes*. Almost every presented prototype was build on data from empirical study(ies) based on problems participants had or reported with PIM (several are discussed at the beginning of most sections). However, while some prototypes tried to solve reported PIM problems others tried to investigate problems in more details. The aim of the latter might not be to be used by users at all. Some prototypes are admittedly not even addressing profound PIM issues (e.g. visualisation prototypes). Many were built just to explore novel interactions and management possibilities with information (e.g. Web book). And some prototypes were build for fun or even art (e.g. Anymail, Mail Garden).

A claim that radical inventions work against technology acquisition (Whittaker et al., 2000) is true to some extent (an example are the above mentioned PIM prototypes for project/task management). Yet, as we will see later, there are many factors that affect adoption and without radical inventions it would certainly be hard to advance any field – PIM included. Due to the slow adoption through gaining a critical mass of researchers and testing, it seems that the vast majority of prototyping efforts never make it into viable products used by significant numbers of people. Nonetheless, we have shown many examples that disprove this. In some cases the causality of transference is unobstructed such as between the SIS search engine prototype (and subsequent Phlat and Milestones in time) developed by Microsoft researchers and search engine in Microsoft Windows. Or even between Treemap algorithm developed in academia and used by many commercial systems of various companies. However, in many cases the causality is not so obvious as we will see later.

So far we only implied one direction of technology transference – from academia to mainstream. The picture of transference complexity is far from this simple. Some technologies were equally popular and studied in both academia and industry at the same time and other technologies developed in industry had a significant impact on the research in academia. For example tagging was qually popular; some early academia prototypes and commercial solutions were developed and lunched during the same period (e.g. MP3 ID tags, BeOS operating system, Presto), while with the explosion of Web 2.0 technologies tagging research and development was equally present in both areas. In such cases it is difficult to discover which development/research affected which.

This is easier when technologies and standards developed in industry open new grounds for research in academia. One such example is the development of open platforms and APIs on which new PIM research prototypes with unexplored possibilities are build (e.g. web browser' add ons, online PIM prototypes mashed with commercial services such as maps, etc.) Another example are standards such as WEBDAV that facilitates collaboration and editing of documents stored on web servers, or iCalendar developed for calendar sharing (although calendars as PIM tools are not discussed in this paper), or a combination of both. These standards influenced further PIM research which was not possible before. In many cases it even seems that the transference causality is turned only the other way around as many PIM academic papers reference and cite commercial products, while unfortunately these links are harder to detect from industry to academia even if they are reciprocal. Nevertheless, new and widely used PIM applications developed in the industry are important for studies of PIM practices and how these change with newly introduced technologies.

Because the transference might not be so clear, it is hard to see the connection between a PIM research prototype and a successful commercial PIM product. For example between one of the earliest email threading in Mona email client (1993) and Gmail, or the Glimpse search engine (1994) and OS X Spotlight, or many prototypes managing thumbnails of information and file thumbnails in file managers. As it has been ascertained by a coherent conclusion of a large body of science and technology studies on history of innovations, small scale field and lab studies cannot guarantee a commercial success (MacKenzie and Wajcman, 1999). The adoption of products is affected by social, economic and market factors.

Findings show that technology is socially shaped in several ways during the whole life-cycle from research to design and usage. This includes among others cultural influences on researchers and designers and selectiveness of potential users and their attempts to find attainable and creative usage of new technologies (Bijker and Law, 1992; Woolgar, 1991). Thus the flexibility of products which are open to creative use can play a crucial role in technology adoption (Haddon, 2005). Besides the social aspect, the commercial success of a product establishes dependency that is hard to break even if better products exist on the market. (Arthur, 1994). Such technology dependencies can play a role in the market dominance of a product until a new disruptive technology is introduced (Anderson and Tushman, 1990).

Frohlich and Sarvas (2011) provide examples of successful and unsuccessful product launches. In their opinion a successful product needs to benefit users, be reliable, have a business value, must be supported by existing technologies and must be designed effectively. This goes hand in hand with our observations of smooth integration into current technologies and practices and industry support. The authors listed four interrelated research grounds that need to be addressed in a potentially successful product innovation:

- User research: to identify potential benefits of a technology and assess their realization in products.
- Technology research: required to develop technical solutions delivering those benefits at low enough costs.
- Design research: required to integrate and present the technology in an attractive and accessible form.
- Business research: needed to position and market the technology to the right group of people at a cost they can afford.

Innovation can be initiated in each of the four grounds but ultimately needs to address all four of them. In authors' opinion the HCI research is addressing only the first three grounds and does not reach far enough to address the business level. The same is certainly true for PIM research prototypes.

Looking at the presented prototypes, the vast majority of them were just this: prototypes meant to investigate PIM usage, design and technology, and were not fully featured applications to support rich everyday PIM activities. Prototypes are often abandoned (or not developed and evaluated further) for several reasons: lack of resources, hard to integrate ideas into current PIM tools and practices (due to closed nature of such tools - e.g. file managers), changed ideas and research interests or simply because the initial study was all the researchers wanted in the first place. All these reasons are very common in academia. Thus, the lack of direct transference to commercial products cannot be seen as a failure if the business research ground has not even been considered in the first place. PIM research prototypes are a clear example of throwaway prototyping which seems to be very common in HCI (Whittaker et al., 2000; Frohlich and Sarvas, 2011). Collaborative development between academia and industry might provide better possibilities for research prototypes down-streaming to commercial products. Looking at the last columns of the tables in Figure 1, collaboration between researchers from academia and industry exists, although less than we would have wished for.

8. CONCLUSION

This paper provides an overview of research prototypes for managing personal information. A glance at the references reveals a great variety of journals, conferences and other publications; the total number of venues where presented ideas, solutions and studies were published is 67. These PIM research prototypes' papers have been published in research areas of artificial intelligence, user modelling, adaptive systems, human computer interaction, information visualisation, information retrieval, computer graphics, visual languages, semantic web, CSCW, web technologies, data and information management, computer vision, database research, software engineering, knowledge management and machine learning. The broad variety of possible venues for PIM research has its attractiveness, however it also makes a comparison of studies harder. Different disciplines have diverse variations of understanding the prototype evaluation and it would be impossible to expect a user study for every PIM prototype as some of them focused on novel ways of interaction, visualisation, underlying systems and algorithms for sorting/organizing, or were simply build as art. However, PIM prototypes developed in other research fields are being increasingly tested with users in the last decade perhaps due to the stricter requirements of conference publishing and to interdisciplinary collaboration.

We have shown that even if individual prototypes are not extensively evaluated and directly transferred to the mainstream, many ideas behind them slowly trickle into our everyday life. Especially ideas that attracted enough research interest and were developed and evaluated through many iterations by many researchers. These include desktop search engines, email threading, filters and rules, enhancing hierarchies with thumbnails, tags, file hierarchy visualisations and many more that found their way into small applications or PIM software extensions. Many such small applications and extensions might not be used by hordes of people, which can be attributed to the personalised nature of PIM (e.g. the plethora of note taking applications), or users simply do not know about them due to the lack of exposure or marketing. Yet they are indispensable to the PIM of users that use them. Some might argue that good ideas make it into the mainstream just because they are good. However, there are probably many good ideas presented here that have not made it yet. We can speculate about the reasons for this and they might again include the lack of resources (people, money, marketing), no noticeable profit or maybe even no interest for enhancing existing PIM tools (e.g. if they have a big market share). This is beyond the scope of the paper. How all these technologies change our PIM behaviour is also beyond this review and we already mentioned some results from the studies that looked into this (Bergman et al., 2008; Civan et al., 2008; Weinreich et al., 2008; Whittaker et al., 2011)

REFERENCES

Abrams, D. and Baecker, R. (1997). How people use www bookmarks. In *CHI '97: Extended abstracts on Human factors in computing systems*, pages 341–342, New York, NY, USA. ACM.

Agarawala, A. and Balakrishnan, R. (2006). Keepin' it real: pushing the desktop metaphor with physics, piles and the pen. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1283–1292, New York, NY, USA. ACM.

Agrawal, R., Roberto J. Bayardo, J., and Srikant, R. (2000). Athena: Mining-based interactive management of text database. In *EDBT '00: Proceedings of the 7th International Conference on Extending Database Technology*, pages 365–379, London, UK. Springer-Verlag.

Amento, B., Terveen, L., Hill, W., and Hix, D. (2000). Topicshop: Enhanced support for evaluating and organizing collections of web sites. In *UIST '00: Proceedings of the ACM symposium on User interface software and technology*, pages 201–209. ACM Press.

Ames, A., Maltzahn, C., Bobb, N., Miller, E. L., Brandt, S. A., Neeman, A., Hiatt, A., and Tuteja, D. (2005). Richer file system metadata using links and attributes. *Proceedings of IEEE / NASA Goddard Conference on Mass Storage Systems and Technologies*, 0:49–60.

Anderson, P. and Tushman, M. L. (1990). Technological discontinuities and dominant designs: A cyclical model of technological change. *Administrative science quarterly*, pages 604–633.

Andrews, K. and Heidegger, H. (1998). Information slices: Visualising and exploring large hierarchies using cascading, semi-circular discs. In *INFOVIS '98: Proceedings of the IEEE Symposium on Information Visualization*, pages 9–11.

Arthur, W. B. (1994). *Increrasing returns and path dependence in the economy*. University of Michigan Press.

Aula, A., Jhaveri, N., and Käki, M. (2005). Information search and re-access strategies of experienced web users. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*, pages 583–592, New York, NY, USA. ACM.

Ayers, E. Z. and Stasko, J. T. (1995). Using graphic history in browsing the world wide web. Technical report, Georgia Institute of Technology.

Baeza-Yates, R., Jones, T., and Rawlins, G. (1996). A new data model: Persistent attribute-centric objects. Technical report, University of Chile.

Bälter, O. and Sidner, C. L. (2002). Bifrost inbox organizer: giving users control over the inbox. In *NordiCHI '02: Proceedings of the second Nordic conference on Human-computer interaction*, pages 111–118, New York, NY, USA. ACM.

Barreau, D. (2008). The persistence of behavior and form in the organization of personal information. *Journal of the American Society for Information Science and Technology (JASIST)*, 59(2):307–317.

Barreau, D. and Nardi, B. (1995). Finding and reminding: file organization from the desktop. *SIGCHI Bulletin*, 27(3):39–43.

Bauer, D., Fastrez, P., and Hollan, J. (2004). Computationally-Enriched 'Piles' for Managing Digital Photo Collections. In *VL/HCC '04: IEEE Symposium on Visual Languages and Human Centric Computing*, pages 193–195.

Bellotti, V., Ducheneaut, N., Howard, M., and Smith, I. (2003). Taking email to task: the design and evaluation of a task management centered email tool. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 345–352, New York, NY, USA. ACM.

Bellotti, V. and Smith, I. (2000). Informing the design of an information management system with iterative fieldwork. In *DIS '00: Proceedings of the 3rd conference on Designing interactive systems*, pages 227–237, New York, NY, USA. ACM.

Bellotti, V. and Thornton, J. (2006). Managing activities with TV-ACTA: Taskvista and activity-centered task assistant. In *SIGIR '06: Personal Information Management Workshop*, volume 4.

Bergman, O., Beyth-Marom, R., and Nachmias, R. (2006). The project fragmentation problem in personal information management. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 271–274, New York, NY, USA. ACM.

Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., and Whittaker, S. (2008). Improved Search Engines and Navigation Preference in Personal Information Management. *ACM Transactions on Information Systems (TOIS)*, 26(4). 2nd International Workshop on Personal Information Management, Seattle, WA, AUG 10-11, 2006.

Bergman, O., Boardman, R., Gwizdka, J., and Jones, W. (2004). Personal information management. In *CHI '04: Extended abstracts on Human factors in computing systems*, pages 1598–1599, New York, NY, USA. ACM.

Bergman, O., Tucker, S., Beyth-Marom, R., Cutrell, E., and Whittaker, S. (2009). It's not that important: demoting personal information of low subjective importance using grayarea. In *CHI '09: Proceedings of the international conference on Human factors in computing systems*, CHI '09, pages 269–278, New York, NY, USA. ACM.

Bijker, W. E. and Law, J. (1992). *Shaping Technology/Building Society: studies in socio-technical change*. MIT press.

Bladh, T., Carr, D. A., and Scholl, J. (2004). Extending tree-maps to three dimensions: A comparative study. Technical report, Luleå University of Technology.

Bloehdorn, S., Görlitz, O., Schenk, S., Völkel, M., and Karlsruhe, F. I. (2006). Tagfs - tag semantics for hierarchical file systems. In *I-KNOW 06: Proceedings of the 6th International Conference on Knowledge Management*, pages 6–8.

Boardman, R. (2004). *Improving tool support for personal information management*. PhD thesis, Imperial college London, University of London.

Boardman, R. and Sasse, M. A. (2004). "stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 583–590, New York, NY, USA. ACM.

Boardman, R., Spence, R., and Sasse, M. A. (2003). Too many hierarchies? The daily struggle for control of the workspace. In *HCI International '03: Proceedings of Interantional Conference on Human-Computer Interaction*.

Bondarenko, O. and Janssen, R. (2005). Documents at hand: Learning from paper to improve digital technologies. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 121–130, New York, NY, USA. ACM.

Boone, G. (1998). Concept features in re:agent, an intelligent email agent. In *AGENTS '98: Proceedings of the second annual conference on Autonomous Agents*, pages 141–148. ACM Press.

Byrne, M. D., John, B. E., Wehrle, N. S., and Crow, D. C. (1999). The tangled web we wove: a taskonomy of www use. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 544–551, New York, NY, USA. ACM.

Card, S., Robertson, G., and Mackinlay, J. (1991). The information visualizer, an information workspace. In *CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 181–186. ACM New York, NY, USA.

Card, S. K., Robertson, G. G., and York, W. (1996). The webbook and the web forager: an information workspace for the world-wide web. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 111–ff., New York, NY, USA. ACM.

Carriere, J. and Kazman, R. (1995). Interacting with huge hierarchies: beyond cone trees. In *IV '95: Proceedings of the Fifth International Conference on Information Visualisation*, pages 74–81. IEEE.

Castro, P. and Lopes, A. (2009). Magnet mail: A visualization system for email information retrieval. In Butz, A., Fisher, B., Christie, M., Kruger, A., Olivier, P., and Theron, R., editors, *Smart Graphics*, volume 5531 of *Lecture Notes in Computer Science*, pages 213–222. Springer Berlin - Heidelberg.

Catarci, T., Dix, A., Katifori, A., Lepouras, G., and Poggi, A. (2007). Task-Centred Information Management. *Digital Libraries: Research and Development*, 4877/2007:197–206.

Chau, D. H., Myers, B., and Faulring, A. (2008). What to do when search fails: finding information by association. In *CHI '08: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 999–1008, New York, NY, USA. ACM.

Chen, J., Guo, H., Wu, W., and Xie, C. (2009). Search your memory ! - an associative memory based desktop search system. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, SIGMOD '09, pages 1099–1102, New York, NY, USA. ACM.

Cheyer, A., Park, J., and Giuli, R. (2005). Iris: Integrate. relate. infer. share. In Decker, S., Park, J., Quan, D., and Sauermann, L., editors, *ISWC '05: Workshop on The Semantic Desktop - Next Generation Personal Information Management and Collaboration Infrastructure at the International Semantic Web Conference*, volume 175.

Chin, R. (2002). Three-dimensional file system browser. *Crossroads*, 9(1):16–18.

Ciccarelli, F. D., Doerks, T., Von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science*, 311(5765):1283–1287.

Civan, A., Jones, W., Klasnja, P., and Bruce, H. (2008). Better to organize personal information by folders or by tags?: The devil is in the details. *ASIST '08: Proceedings of Annual Meeting of the American Society for Information Science and Technology*, 45(1):1–13.

Cockburn, A., Greenberg, S., McKenzie, B., Jasonsmith, M., and Kaasten, S. (1999). WebView: A graphical aid for revisiting Web pages. In *OZCHI '99: Proceedings of the Australia conference on Human-Computer Interaction*, volume 99, pages 15–22.

Cockburn, A. and McKenzie, B. (2001a). 3D or not 3D?: evaluating the effect of the third dimension in a document management system. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, page 441. ACM.

Cockburn, A. and McKenzie, B. (2001b). What do web users do? an empirical analysis of web use. *International Journal of Human Computer Studies*, 54(6):903–922.

Cockburn, A. and Thimbleby, H. (1993). Reducing user effort in collaboration support. In *IUI '93: Proceedings of the 1st international conference on Intelligent user interfaces*, pages 215–218, New York, NY, USA. ACM.

Cutrell, E., Robbins, D., Dumais, S., and Sarin, R. (2006). Fast, flexible filtering with phlat. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 261–270, New York, NY, USA. ACM.

Denning, P. J. (1982). Acm president's letter: electronic junk. *Communications of the ACM*, 25(3):163–165.

Diep, E. and Jacob, R. (2004). Visualizing E-mail with a Semantically Zoomable Interface. In *INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization*.

Dömel, P. (1995). Webmap: a graphical hypertext navigation tool. *Computer Networks and ISDN Systems*, 28(1-2):85–97.

Dong, X. (2005). A platform for personal information management and integration. In *VLDB '05: Workshop at the International Conference on Very Large Databases*, page 26.

Dourish, P., Edwards, W. K., LaMarca, A., and Salisbury, M. (1999). Presto: an experimental architecture for fluid interactive document spaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 6(2):133–161.

Dragunov, A., Dietterich, T., Johnsrude, K., McLaughlin, M., Li, L., and Herlocker, J. (2005). TaskTracer: a desktop environment to support multi-tasking knowledge workers. In *IUI '05: Proceedings of the 10th international conference on Intelligent user interfaces*, pages 75–82. ACM.

Ducheneaut, N. and Bellotti, V. (2001). E-mail as habitat: an exploration of embedded personal information management. *Interactions Magazine*, 8(5):30–38.

Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., and Robbins, D. C. (2003). Stuff i've seen: a system for personal information retrieval and re-use. In *SIGIR '03: Proceedings of the ACM SIGIR conference on Research and development in information retrieval*, pages 72–79. ACM Press.

Dumais, S., Cutrell, E., and Chen, H. (2001). Optimizing search by showing results in context. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, page 284. ACM.

Dumais, S., Cutrell, E., Sarin, R., and Horvitz, E. (2004). Implicit queries (IQ) for contextualized search. In *SIGIR '04: Proceedings of the ACM SIGIR conference on Research and development in information retrieval*, page 594. ACM.

Espenkrona, K. and Svensson, M. (2009). Composing an evaluation framework for personal information management tools. Bachelor Thesis Report No. 2009:043, University of Gothenburg, Department of Applied Information Technology.

Farnham, S., Portnoy, W., Turski, A., Cheng, L., and Vronay, D. (2003). Personal map: Automatically modeling the user's online social network. In *INTERACT '03: Proceedings of IFIP TC13 Conference on Human-Computer Interaction*, pages 567–574.

Fertig, S., Freeman, E., and Gelernter, D. (1996). "'Finding and reminding"' reconsidered. *SIGCHI Bulletin*, 28(1):66–69.

Fisher, D., Brush, A. J., Gleave, E., and Smith, M. A. (2006). Revisiting Whittaker & Sidner's "'email overload"' ten years later. In *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 309–312, New York, NY, USA. ACM.

Frau, S., Roberts, J. C., and Boukhelifa, N. (2005). Dynamic coordinated email visualization. In Skala, V., editor, *WSCG '05: International Conference on Computer Graphics, Visualization and Computer Vision*, pages 187–193, Plzen, Czech Republic. (Jan 31 - Feb 4).

Frecon, E. and Smith, G. (1998). Webpath - a three dimensional web history. *INFOVIS '88: Proceedings of the IEEE Symposium on Information Visualization*, pages 3 – 10, 148.

Freeman, E. and Gelernter, D. (1996). Lifestreams: a storage model for personal data. *SIGMOD Record*, 25(1):80–86.

Frohlich, D. M. and Sarvas, R. (2011). Hci and innovation. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 713–728. ACM.

Gandhi, R., Kumar, G., Bederson, B., and Shneiderman, B. (2000). Domain name based visualization of web histories in a zoomable user interface. *DEXA '00: Proceedings of International Workshop on Database and Expert Systems Applications*, pages 591 –598.

Gifford, D., Jouvelot, P., Sheldon, M., et al. (1991). Semantic file systems. *ACM SIGOPS Operating Systems Review*, 25(5):16–25.

Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70.

Gomes, P., Gama, S., and Gonçalves, D. (2010). Designing a personal information visualization tool. In *NordiCHI '10; Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 663–666, New York, NY, USA. ACM.

Gonçalves, D. and Jorge, J. A. (2008). In search of personal information: narrative-based interfaces. In *IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces*, pages 179–188, New York, NY, USA. ACM.

Grivet, S., Auber, D., Domenger, J.-P., and Melancon, G. (2006). Bubble tree drawing algorithm. In *Computer Vision and Graphics*, pages 633–641. Springer.

Groza, T., Handschuh, S., Moeller, K., Grimnes, G., Sauermann, L., Minack, E., Mesnage, C., Jazayeri, M., Reif, G., and Gudjonsdottir, R. (2007). The nepomuk project-on the way to the social semantic desktop. *I-SEMANTICS '07: Proceedings of International Conference on Semantic Technologies*, 7:201–211.

Gwizdka, J. (2002). Taskview: design and evaluation of a task-based email interface. In *CASCON '02: Proceedings of the 2002 conference of the Centre for Advanced Studies on Collaborative research*, page 4. IBM Press.

Haddon, L. (2005). *Everyday Innovators: Researching the role of users in shaping ICTs*, volume 32. Springer.

Hailpern, J., Jitkoff, N., Warr, A., Karahalios, K., Sesek, R., and Shkrob, N. (2011). Youpivot: improving recall with contextual search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1521–1530, New York, NY, USA. ACM.

Hascoët, M. (2000). A user interface combining navigation aids. In *HYPERTEXT '00: Proceedings of the eleventh ACM on Hypertext and hypermedia*, pages 224–225, New York, NY, USA. ACM.

Hearst, M. A. and Karadi, C. (1997). Cat-a-cone: an interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. In *ACM SIGIR Forum*, volume 31, pages 246–255. ACM.

Henderson, J. D. A. and Card, S. (1986). Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics (TOG)*, 5(3):211–243.

Hightower, R. R., Ring, L. T., Helfman, J. I., Bederson, B. B., and Hollan, J. D. (1998). Graphical multiscale web histories: a study of padprints. In *HYPERTEXT '98: Proceedings of the ninth ACM conference on Hypertext and hypermedia*, pages 58–65, New York, NY, USA. ACM.

Holten, D. (2006). Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):741–748.

Horn, C. (2007). Natural metaphor for information visuzalization. Master's thesis, Massachusetts College of Art in Boston.

Jensen, C., Lonsdale, H., Wynn, E., Cao, J., Slater, M., and Dietterich, T. G. (2010). The life and times of files and information: a study of desktop provenance. In *CHI '10: Proceedings of the international conference on Human factors in computing systems*, CHI '10, pages 767–776, New York, NY, USA. ACM.

Jeong, C.-S. and Pang, A. (1998). Reconfigurable disc trees for visualizing large hierarchical information space. In *IV '98: Proceedings of the International Conference on Information Visualisation*, pages 19–25. IEEE.

Johnson, B. and Shneiderman, B. (1991). Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In *Visualization '91: Proceedings of the 2nd IEEE conference on Visualization*, pages 284–291. IEEE Computer Society Press.

Jones, W. (2007). Personal information management. *Annual review of information science and technology*, 41(1):453–504.

Jones, W., Bruce, H., and Dumais, S. (2001). Keeping found things found on the web. In *CIKM '01: Proceedings of the tenth international conference on Information and knowledge management*, pages 119–126, New York, NY, USA. ACM.

Jones, W., Dumais, S., and Bruce, H. (2002). Once found, what then? A study of "keeping" behaviors in the personal use of Web information. *ASIST '02: Proceedings of Annual Meeting of the American Society for Information Science and Technology*, 39(1):391–402.

Jones, W., Hou, D., Sethanandha, B. D., Bi, S., and Gemmell, J. (2010). Planz to put our digital information in its place. In *CHI '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, CHI EA '10, pages 2803–2812, New York, NY, USA. ACM.

Jones, W., Klasnja, P., Civan, A., and Adcock, M. L. (2008). The personal project planner: planning to organize personal information. In *CHI '08: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 681–684, New York, NY, USA. ACM.

Jones, W., Munat, C., and Bruce, H. (2005). The universal labeler: Plan the project and let your information follow. In *ASIST '05: Proceedings of Annual Meeting of the American Society for Information Science and Technology.* American Society for Information Science and Technology.

Kaasten, S. and Greenberg, S. (2001). Integrating back, history and bookmarks in web browsers. In *CHI '01: Extended abstracts on Human factors in computing systems*, pages 379–380, New York, NY, USA. ACM.

Kaptelinin, V. (2003). Umea: translating interaction histories into project contexts. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 353–360, New York, NY, USA. ACM.

Karger, D. (2007). Unify everything: it's all the same to me. In *Personal Information Management*, chapter 8, pages 127–152. Univ of Washington Pr.

Karger, D. R., Bakshi, K., Huynh, D., Quan, D., and Sinha, V. (2005). Haystack: A customizable general-purpose information management tool for end users of semistructured data. In *CIDR '05: Proceedings of the Conference on Innovative Database Research*, pages 13–26.

Karger, D. R. and Jones, W. (2006). Data unification in personal information management. *Communications of the ACM*, 49(1):77–82.

Karlson, A., Smith, G., and Lee, B. (2011). Which version is this?: improving the desktop experience within a copy-aware computing ecosystem. In *CHI '11: Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 2669–2678. ACM.

Kelley, D. and Teevan, J. (2007). Understanding what works: Evaluating pim tools. In Jones, W. and Teevan, J., editors, *Personal Information Management*. University of Washington Press, Seattle, WA.

Kerr, B. (2003). Thread arcs: An email thread visualization. *INFOVIS '03: Proceedings of the IEEE Symposium on Information Visualization*, 0:27.

Kerr, B. and Wilcox, E. (2004). Designing remail: reinventing the email client through innovation and integration. In *CHI '04: Extended abstracts on Human factors in computing systems*, pages 837–852, New York, NY, USA. ACM.

Killam, B. (2001). A study of three browser history mechanisms for web navigation. In *IV '01: Proceedings of the Fifth International Conference on Information Visualisation*, Washington, DC, USA. IEEE Computer Society.

Kleiberg, E., Van De Wetering, H., and Van Wijk, J. J. (2001). Botanical visualization of huge hierarchies. In *INFOVIS '01: Proceedings of the IEEE Symposium on Information Visualization*, volume 1, page 87.

Kobsa, A. (2004). User experiments with tree visualization systems. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 9–16. IEEE.

Koike, H. and Yoshihara, H. (1993). Fractal approaches for visualizing huge hierarchies. In *VL '93: Proceedings of the IEEE Symposium on Visual Languages*, pages 55–60. IEEE.

Kwasnik, B. (1989). *The influence of context on classification behavior*. PhD thesis, Rutgers University, East Rutherford, NY.

Lam, B. (2005). Mailstacker: Applying a pile metaphor for email management. Bachelors Theses, School of Information Technology and Electrical Engineering, The University of Queensland.

Lampasona, C., Rostanin, O., and Maus, H. (2012). Seamless integration of order processing in ms outlook using smartoffice: an empirical evaluation. In *ESEM '12: Proceedings of the International symposium on Empirical software engineering and measurement*, pages 165–168. ACM.

Lamping, J., Rao, R., and Pirolli, P. (1995). A focus+ context technique based on hyperbolic geometry for visualizing large hierarchies. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 401–408. ACM Press/Addison-Wesley Publishing Co.

Lansdale, M. and Edmonds, E. (1992). Using memory for events in the design of personal filing systems. *International Journal of Man-Machine Studies*, 36(1):97–126.

Lepouras, G., Papatriantafyllou, A., Katifori, A., and Dix, A. (2008). Time2hide: spatial searches and clutter alleviation for the desktop. In *AVI '08: Proceedings of the working conference on Advanced visual interfaces*, pages 355–358, New York, NY, USA. ACM.

Liao, H., Osada, M., and Shneiderman, B. (1998). Browsing Unix directories with Dynamic Queries: An evaluation of three information display techniques. Technical Report CS-TR-2841, UM Computer Science Department.

Mackay, W. E. (1988). Diversity in the use of electronic mail: a preliminary inquiry. *ACM Transactions on Information Systems (TOIS)*, 6(4):380–397.

MacKenzie, D. and Wajcman, J. (1999). *The social shaping of technology*. Open University Press.

Mackinlay, J., Robertson, G., and Card, S. (1991). The perspective wall: Detail and context smoothly integrated. In *CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 173–176. ACM.

Madlock-Brown, C., Chin, S.-C., and Eichmann, D. (2012). Infoatv: enhancing pim systems to support re-finding to (re)discovery. In *CSCW '12: Workshop on Personal Information Management*, volume 26.

Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM*, 37(7):30–40.

Malone, T. W. (1983). How do people organize their desks?: Implications for the design of office information systems. *ACM Transactions on Information Systems (TOIS)*, 1(1):99–112.

Malone, T. W., Grant, K. R., Turbak, F. A., Brobst, S. A., and Cohen, M. D. (1987). Intelligent information-sharing systems. *Communications of the ACM*, 30(5):390–402.

Manber, U. and Wu, S. (1994). Glimpse: a tool to search through entire file systems. In *WTEC'94: Proceedings of the USENIX Winter 1994 Technical Conference*, pages 4–4, Berkeley, CA, USA. USENIX Association.

Mander, R., Salomon, G., and Wong, Y. Y. (1992). A "pile" metaphor for supporting casual organization of information. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 627–634, New York, NY, USA. ACM.

Mandic, M. and Kerne, A. (2004). famailiar & intimacy-based email visualization. *INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization*, pages p14 – p14.

Marsden, G. and Cairns, D. (2003). Improving the usability of the hierarchical file system. In *SAICSIT '03: Proceedings of the research conference of the South African institute of computer scientists and information technologists*, page 129. South African Institute for Computer Scientists and Information Technologists.

Mayer, M. and Bederson, B. (2003). Browsing Icons: A Task-Based Approach for a Visual Web History. Technical report, UM Computer Science Department.

Moore, B., Van Kleek, M., Karger, D. R., and Schraefel,

M. (2010). Assisted self reflection: Combining lifetracking, sensemaking, & personal information management. In *CHI 2010: Workshop - Know Thyself: Monitoring and Reflecting on Facets of One's Life*.

Morteo, R., González, V. M., Favela, J., and Mark, G. (2004). Sphere juggler: Fast context retrieval in support of working spheres. In *Proceedings of the Fifth Mexican International Conference in Computer Science*, ENC '04, pages 361–367, Washington, DC, USA. IEEE Computer Society.

Muller, M., Geyer, W., Brownholtz, B., Wilcox, E., and Millen, D. (2004). One-hundred days in an activity-centric collaboration environment based on shared objects. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 375–382. ACM.

Muniswamy-Reddy, K., Holland, D., Braun, U., and Seltzer, M. (2006). Provenance-aware storage systems. In *USENIX '06: Proceedings of the 2006 USENIX Annual Technical Conference*, pages 43–56.

Nardi, B., Miller, J., and Wright, D. (1998). Collaborative, programmable intelligent agents. *Communications of the ACM*, 41(3):96–104.

Neumann, P., Carpendale, M. S. T., and Agarawala, A. (2006). Phyllotrees: Phyllotactic patterns for tree layout. In *EuroVis '06: Proceedings of Eurographics / IEEE VGTC Symposium on Visualization*, pages 59–66. Eurographics.

Padioleau, Y. and Ridoux, O. (2002). A Logic File System. Research Report RR-4656, INRIA.

Pedersen, E. and McDonald, D. (2008). Relating documents via user activity: the missing link. In *IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces*, pages 389–392. ACM.

Perer, A., Shneiderman, B., and Oard, D. W. (2006). Using rhythms of relationships to understand e-mail archives. *Journal of the American Society for Information Science and Technology (JASIST)*, 57(14):1936–1948.

Perer, A. and Smith, M. A. (2006). Contrasting portraits of email practices: visual approaches to reflection and analysis. In *AVI '06: Proceedings of the working conference on Advanced visual interfaces*, pages 389–395, New York, NY, USA. ACM.

Quan, D., Bakshi, K., Huynh, D., and Karger, D. (2003). User interfaces for supporting multiple categorization. In *INTERACT '03: Proceedings of the Ninth IFIP TC13 International Conference on Human-Computer Interaction*, pages 228 – 236.

Rattenbury, T. and Canny, J. (2007). Caad: an automatic task support system. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 687–696. ACM.

Ravasio, P., Schr, S. G., and Krueger, H. (2004). In pursuit of desktop evolution: User problems and practices with modern desktop systems. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 11(2):156–180.

Ravasio, P., Vukelja, L., Rivera, G., and Norrie, M. (2003). Project infospace: From information managing to information representation. In *INTERACT '03: Proceedings of IFIP TC13 Conference on Human-Computer Interaction*.

Rekimoto, J. (1999). Time-machine computing: a time-centric approach for the information environment. In *UIST '99: Proceedings of the ACM symposium on User interface software and technology*, pages 45–54, New York, NY, USA. ACM.

Rekimoto, J. and Green, M. (1993). The information cube: Using transparency in 3d information visualization. In *WITS '93: Proceedings of the Workshop on Information Technologies & Systems*, pages 125–132.

Rhodes, B. and Starner, T. (1996). Remembrance Agent: A continuously running automated information retrieval system. In *PAAMS '96: Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi Agent Technology*, pages 487–495.

Richter, J., V
"olkel, M., and Haller, H. (2005.). Deepamehta - a semantic desktop. In Decker, S., Park, J., Quan, D., and Sauermann, L., editors, *ISWC '05: Workshop on The Semantic Desktop - Next Generation Personal Information Management and Collaboration Infrastructure at the International Semantic Web Conference*, volume 175.

Ringel, M., Cutrell, E., Dumais, S., and Horvitz, E. (2003). Milestones in time: The value of landmarks in retrieving information from personal stores. In *INTERACT '03: Proceedings of IFIP TC13 Conference on Human-Computer Interaction*, pages 184–191.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., and van Dantzich, M. (1998). Data mountain: using spatial memory for document management. In *UIST '98: Proceedings of the ACM symposium on User interface software and technology*, pages 153–162, New York, NY, USA. ACM.

Rohall, S. and Gruen, D. (2002). Remail: A reinvented email prototype. In *CSCW '02: Demonstration on Computer Supported Cooperative Work*, New Orleans, LA.

Rohall, S., Gruen, D., Moody, P., and Kellerman, S. (2001). Email visualizations to aid communications. In *INFOVIS '01: Proceedings of the IEEE Symposium on Information Visualization*.

Rompa, J., Lepouras, G., Vassilakis, C., and Tryfonopoulos, C. (2011). Ontofm: A personal ontology-based file manager for the desktop. In *ISCW '11: Demo at the 10th International Semantic Web Conference*, Bonn, Germany.

Samiei, M., Dill, J., and Kirkpatrick, A. (2004). Ezmail: Using information visualization techniques to help manage email. *IV '04: Proceedings of International Conference on Information Visualisation*, 0:477–482.

Sauermann, L., Dengel, A., Van Elst, L., Lauer, A., Maus, H., and Schwarz, S. (2006a). Personalization in the epos project. In *Proceedings of the International Workshop on Semantic Web Personalization, Budva, Montenegro*, pages 42–52.

Sauermann, L., Grimnes, G., Kiesel, M., Fluit, C., Maus, H., Heim, D., Nadeem, D., Horak, B., and Dengel, A. (2006b). Semantic desktop 2.0: The gnowsis experience. *ISWC '06: Proceedings of International Semantic Web Conference*, pages 887–900.

Schraefel, M., Karam, M., and Zhao, S. (2003). mSpace: interaction design for user-determined, adaptable domain exploration in hypermedia. In *AH '03: Proceedings of the Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems*.

Sebrechts, M. M., Cugini, J. V., Laskowski, S. J., Vasilakis, J., and Miller, M. S. (1999). Visualization of search results: a comparative evaluation of text, 2d, and 3d interfaces. In *SIGIR '99: Proceedings of the ACM SIGIR conference on Research and development in information retrieval*, pages 3–10, New York, NY, USA. ACM.

Segal, R. and Kephart, J. O. (2000). Incremental learning in swiftfile. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, pages 863–870, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Segal, R. B. and Kephart, J. O. (1999). Mailcat: an intelligent assistant for organizing e-mail. In *AGENTS '99: Proceedings of the third annual conference on Autonomous Agents*, pages 276–282, New York, NY, USA. ACM.

Shah, S., Soules, C. A. N., Ganger, G. R., and Noble, B. D. (2007). Using provenance to aid in personal file search. In *USENIX '07: Proceedings of the 2006 USENIX Annual Technical Conference*, ATC'07, pages 13:1–13:14, Berkeley, CA, USA. USENIX Association.

Shen, J., Li, L., Dietterich, T. G., and Herlocker, J. L. (2006). A hybrid learning system for recognizing user tasks from desktop activities and email messages. In *IUI '06: Proceedings of the 11th international conference on Intelligent user interfaces*, pages 86–92, New York, NY, USA. ACM.

Shirai, Y., Yamamoto, Y., and Nakakoji, K. (2006). A history-centric approach for enhancing web browsing experiences. In *CHI '06: Extended abstracts on Human factors in computing systems*, pages 1319–1324, New York, NY, USA. ACM.

Shneiderman, B. and Plaisant, C. (1994). The future of graphic user interfaces: personal role managers. *People and Computers*, pages 3–9.

Sinha, V. and Karger, D. R. (2005). Magnet: supporting navigation in semistructured data environments. In *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 97–106, New York, NY, USA. ACM.

Sorensen, C., Macklin, D., and Beaumont, T. (2001). Navigating the world wide web: bookmark maintenance architectures. *Interacting with Computers*, 13(3):375 – 400.

Soules, C. and Ganger, G. (2005). Connections: using context to enhance file search. *ACM SIGOPS Operating Systems Review*, 39(5):119–132.

Stasko, J., Catrambone, R., Guzdial, M., and McDonald, K. (2000). An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies*, 53(5):663–694.

Sudarsky, S. and Hjelsvold, R. (2002). Visualizing electronic mail. In *IV '02: Proceedings of Sixth International Conference on Information Visualisation*, pages 3 – 9.

Takkinen, J. and Shahmehri, N. (1998). Cafe: A conceptual model for managing information in electronic mail. In *HICSS '98: Proceedings of the Thirty-First Annual Hawaii International Conference on System Sciences*, volume 5, page 44, Washington, DC, USA. IEEE Computer Society.

Tang, J. C., Wilcox, E., Cerruti, J. A., Badenes, H., Nusser, S., and Schoudt, J. (2008). Tag-it, snag-it, or bag-it: combining tags, threads, and folders in e-mail. In *CHI '08 extended abstracts on Human factors in computing systems*, CHI EA '08, pages 2179–2194, New York, NY, USA. ACM.

Tauscher, L. and Greenberg, S. (1997). How people revisit web pages: empirical findings and implications for the design of history systems. *International Journal of Human Computer Studies*, 47:97–138.

Teevan, J., Cutrell, E., Fisher, D., Drucker, S., Ramos, G., André, P., and Hu, C. (2009). Visual snippets: summarizing web pages for search and revisitation. In *CHI '09: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 2023–2032. ACM New York, NY, USA.

Teraoka, T. and Maruyama, M. (1997). Adaptive information visualization based on the user's multiple viewpoints-interactive 3d visualization of the www. In *IV '97: Proceedings of the International Conference on Information Visualisation*, pages 25–28. IEEE.

Tulving, E. and Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5(4):381 – 391.

Tungare, M. and Pérez-Quiñones, M. A. (2009). Mental workload in multi-device personal information management. In *ASIST '09: Personal Information Management (PIM) Workshop*.

Van Ham, F. and van Wijk, J. J. (2003). Beamtrees: Compact visualization of large hierarchies. *Information Visualization*, 2(1):31–39.

van Wijk, J., van Ham, F., and van de Wetering, H. (2003). Rendering hierarchical data. *Communications of the ACM*, 46(9):263.

Van Wijk, J. J. and Van de Wetering, H. (1999). Cushion treemaps: Visualization of hierarchical information. In *INFOVIS '99: Proceedings of the IEEE Symposium on Information Visualization*, pages 73–78. IEEE.

Venolia, G., Dabbish, L., Cadiz, J., and Gupta, A. (2001). Supporting email workflow. Technical Report MSR-TR-2001-88, Microsoft Research.

Viegas, F. (2005). Mountain. http://alumni.media.mit.edu/{~}fviegas/projects/mountain/index.htm.

Viégas, F., Boyd, D., Nguyen, D., Potter, J., and Donath, J. (2004). Digital artifacts for remembering and storytelling: Posthistory and social network fragments. In *HICSS '04: Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, page 10.

Viégas, F. B., Golder, S., and Donath, J. (2006). Visualizing email content: portraying relationships from conversational histories. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 979–988, New York, NY, USA. ACM.

Voida, S. and Mynatt, E. (2009). It feels better than filing: everyday work experiences in an activity-based computing system. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 259–268. ACM.

Voit, K., Andrews, K., and Slany, W. (2009). Why Personal Information Management (PIM) Technologies Are Not Widespread. In *ASIST '09: Personal Information Management (PIM) Workshop*, Vancouver, BC, Canada. ASIS&T 2009 Workshop.

Voit, K., Andrews, K., and Slany, W. (2011). tagstore and tagtree: Storing and re-finding files using tags. BCS HCI '11: Poster on the British HCI Group Annual Conference.

Weinreich, H., Obendorf, H., Herder, E., and Mayer, M. (2006). Off the beaten tracks: exploring three aspects of web navigation. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 133–142, New York, NY, USA. ACM.

Weinreich, H., Obendorf, H., Herder, E., and Mayer, M. (2008). Not quite the average: An empirical study of web use. *ACM Trans. Web*, 2(1):5:1–5:31.

Wexelblat, A. and Maes, P. (1999). Footprints: history-rich tools for information foraging. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 270–277, New York, NY, USA. ACM.

Whittaker, S. (2011). Personal information management: from information consumption to curation. *Annual Review of Information Science and Technology*, 45:3–62.

Whittaker, S., Jones, Q., Nardi, B., Creech, M., Terveen, L., Isaacs, E., and Hainsworth, J. (2004). Contactmap: Organizing communication in a social desktop. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 11(4):445–471.

Whittaker, S., Matthews, T., Cerruti, J., Badenes, H., and Tang, J. (2011). Am i wasting my time organizing email?: a study of email refinding. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 3449–3458, New York, NY, USA. ACM.

Whittaker, S. and Sidner, C. (1996). Email overload: exploring personal information management of email. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 276–283, New York, NY, USA. ACM.

Whittaker, S., Terveen, L., and Nardi, B. A. (2000). Let's stop pushing the envelope and start addressing it: a reference task agenda for hci. *Hum.-Comput. Interact.*, 15(2):75–106.

Wikipedia (2005). Archy - the humane environment project. http://en.wikipedia.org/wiki/Archy. The Archy name was chosen on Jan 1 2005.

Wilkens, K. (2007). Mail garden. http://human-centered-visualizations.com/.

Wolber, D., Kepe, M., and Ranitovic, I. (2002). Exposing document context in the personal web. In *IUI '02: Proceedings of the 7th international conference on Intelligent user interfaces*, pages 151–158, New York, NY, USA. ACM.

Won, S. S., Jin, J., and Hong, J. I. (2009). Contextual web history: using visual and contextual cues to improve web browser history. In *CHI '09: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1457–1466, New York, NY, USA. ACM.

Woolgar, S. (1991). *Configuring the User: The Case of Usability Trials.*, chapter 3, pages 57–99. Routledge, New York.

Xu, W., Esteva, M., and Jain, S. (2010). Visualizing personal digital collections. In *JCDL '10: Proceedings of the 10th annual joint conference on Digital libraries*, pages 169–172. ACM.

Yamaguchi, T., Hattori, H., Ito, T., and Shintani, T. (2004). On a web browsing support system with 3d visualization. In *WWW Alt. '04: Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, pages 316–317, New York, NY, USA. ACM.

Yee, K. (2002). Zest: discussion mapping for mailing lists. In *CSCW '02: Proceedings on the ACM Conference on Computer Supported Cooperative Work*, New Orleans, Louisiana, USA. ACM.

Yee, K. P. and Hearst, M. (2005). A visualization to facilitate productive discussions. In *CHI '05: Beyond Threaded Conversation Workshop on SIGCHI conference on Human factors in computing systems.*

Yiu, K. S., Baecker, R., Silver, N., and Long, B. (1997). A time-based interface for electronic mail and task management. *Advances in human factors/ergonomics*, pages 19–22.

Yu, W. and Ingalls, T. (2011). Trails: An interactive web history visualization and tagging tool. In Marcus, A., editor, *DUXU '11: Design, User Experience, and Usability. Theory, Methods, Tools and Practice*, volume 6770 of *Lecture Notes in Computer Science*, pages 77–86. Springer Berlin Heidelberg.

Zhao, L., Deng, T., Wang, H., Liu, Q., and Feng, L. (2012). Design and implementation of a context-based media retrieval system. In Hu, S.-M. and Martin, R., editors, *CVM '12: Computational Visual Media*, volume 7633 of *Lecture Notes in Computer Science*, pages 218–225. Springer Berlin - Heidelberg.