# A Study of a Crosstool Information Usage on Personal Computers: how users mentally link information relating to a task but residing in different applications and how importance and type of acquisition affect this

Matjaz Kljun
m.kljun@lancaster.ac.uk*

Alan Dix
dixa@comp.lancs.ac.uk *

Franc Solina
franc@fri.uni-lj.si †

November 2009

## Abstract

*Our information space is partitioned between real and digital worlds. Information in a digital world is fragmented between several devices such us mobile phones, handheld devices, netbooks, notebooks and personal computers. Even on each device information is fragmented because of several information formats or types (files, emails, web bookmarks. etc.) and applications. To some extent we already understand how we manage documents in the real world, mobile devices and software applications. But we have yet to understand how information between applications, devices and the real world is connected and how we cope with the burden of memorizing connections while we constantly create new ones. This research emphasized information management on personal computers. Our aim was to find out how much information users consider important in three main hierarchy based structures (files, emails and web bookmarks), how information between hierarchies (and applications) is mentally linked and managed according to tasks, how long it is regarded as important and if this information receives any special treatment. An online questionnaire was developed and participants were invited to daily enter the data about information they considered important and to link that information to other information where they thought the link is necessary. The results showed that participants regarded their information as important but the time this information was considered important was short. It also showed that more created information was regarded as important than received or found, that information was mentally linked in the same hierarchy as well as across hierarchies or tools and also that information can be part of several task information collections or that collections overlap. Although this study showed clear evidence that users do mentally link their information, it also raised several question like how links are maintained, how they relate to tasks and how they change over time.*

*InfoLab21, Lancatser University
†Faculty of Computer and Information Science, University of Ljubljana

# 1 Introduction

We all collect different objects such us postcards, magazines, documents, books and letters which fill our shelves, drawers and tables. We act similarly in digital environments. With cheap digital space more and more information is available in digital form. The technology has enabled us to easily create, receive, find, file, modify, copy, move and delete digital information in a more convenient way than we are able to in the real world. But we manage all this information in several applications. Digital documents, photos, music, video and other files can be either created, received (via email) or downloaded (from Web). These files are typically stored in a file system on a Desktop and home folder. Most of these files can be (besides in file manager) managed in specialized photo, music and video managers. Our emails are piling up in a separate hierarchy managed by email client(s) and we might even have several email accounts each resulting a in separate hierarchy. The Web represents a vast collection of documents that we can bookmark in another hierarchy. Then we have calendars, sticky notes, to-do lists, instant messaging (IM) conversations, several contact books (email addresses, IM friends, telephone numbers, post addresses and similar contact information of people we know) and other digital information. Information is managed separately in several applications based on a format or type, even if it is semantically similar and as such would naturally be stored together [24]. This causes *information fragmentation*.

With a vast amount of information scattered and managed across several applications, that have not changed much in last 30 years, we are faced with information overflow [27]. Recent solutions are just small steps towards helping users in managing information. These solutions include addition of icons to files and applications, file thumbnails, different folder views, temporal or virtual folders, email conversation threads, website icons and similar graphical extras. But all these help us only in each application separately. Hierarchies (as a main storage structure) remained static, items locations unique and metadata easily forgettable. Other problems seen in present hierarchies include [15]:

- Organizational structure cannot easily change over time.
- Dynamically organized collections (according to projects, time, content, etc.) are not possible.
- (User controlled) automation in organization is lacking.

An even bigger problem remains of managing information across applications. This has recently resulted in cross tools search engines that index all information in real time. But some researchers have come to the conclusion that users prefer browsing hierarchies step by step (orienteering) over using search engines (teleporting) [26]. One reason for this is that while browsing, users explore and memorize their information space. However, these new and more reliable search engines may change future users' behavior [9].

A term that describes managing information in the real and in digital world(s) is Personal Information Management (PIM). PIM in a digital world is described in detail in the next section. Hypotheses on how users assess their information as important or unimportant, and how information fragmentation affects users' practices are presented in Section 3. Our study of information importance and how information is mentally linked across applications is described in Sections 4.2 and 4.3. At the end we take a look on possible further work in this field.

# 2 Personal Information Management

To understand the importance of information and how information enters in a digital world, some terms should be explained as we will use them throughout this paper. Description of (personal) information, what is a unit of information, how units relate and form more complex structures and what is PIM are presented in next subsection. Divisions of PIM in sub activities are discussed in subsection 2.2.

## 2.1 Definition of PIM

The following are descriptions of terms by different authors that are used to define a framework for Personal Information Management. These terms are: *Personal Information*, *Information Item*, *Personal Information Collection*, *Personal Space of Information* and *Personal Information Management*.

Personal information

Boardman defines *personal information* as information owned by an individual and over which this individual has a direct control [10, p. 14]. Jones and Teevan [17, p. 7] state that *personal information* can take various senses: (1) the information a person directly or indirectly keeps for personal use, (2) information about a person kept under control of others (e.g. medical records), (3) information experienced by a person but not in a person's control (a book from library) and (4) information directed to a person (email).

When mentioning personal information in this paper, we will refer to information a person keeps and information directed to a person in a digital world on personal computer.

Information item

A self contained unit of information is called an *information item* which in digital world exists in different technological formats (files, emails, web pages, etc.) and may contain metadata defined by the user or system (size, creation date, name, etc.) [10, p. 15]. Another definition describes *information item* as *packaging of information in a persistent form that can be acquired, created, moved, grouped with other information items, given name and properties, distributed, deleted or otherwise manipulated. Information items* exists in various *information forms or types* determined by tools or applications that make it possible to manipulate them [17, p. 10].

Personal Information Collection

*Personal Information Collections* (PIC) are formed by information items. But Jones' and Teevan's view of PIC differs from Boarman's view. Boardman describes a *collection of personal information* as a self-contained set of *information items* that share the same technological format and can be accessed in a particular application (with the exception of a file system that holds files in different technological formats): email collection, web bookmark collection, a collection of email addresses, etc [10, p. 15]. On the other hand, Jones and Teevan do not limit PICs to one technological format, rather PICs are formed based on person's activities. As such a PIC can take forms of a sub-collection of well organized files or stored bookmarks or even contain files, emails and other information types. PICs are important as they represent smaller chunks (e.g. folders in a hierarchy) out of PSI which can be managed together [17, p. 11].

When discussing PICs in this paper, we will always refer to later definition.

Personal Space of Information

Boardman calls a complete set of collections *Personal Information Environment* (PIE) [10, p. 26], while Jones and Teevan call it a *Personal Space of Information* (PSI) [17, p. 9]. PSI or PIE is not limited to a digital world. Rather it includes all (four groups of) personal information from the real and digital worlds on different devices (see the above description of Personal Information).

When referring to PSI in this paper, we will always refer to PSI limited to a personal computer.

Personal Information Management

PIM can be seen as managing information items to form Personal Information Collections and managing PICs to form Personal Space of Information. Over the years several definitions of PIM have emerged. Boardman [10, p. 16] defines it as *the management of personal information as performed by the owning individual*. PIM in its basic form is often viewed as a collection of information we manage (handle, categorize) for later retrieval [21]. Some researchers have added that we also store information on visually exposed places such us desktops or email inboxes to remind us of work we have yet to do [24]. This behavior is present also in the real world where we have sticky notes and documents on exposed places to serve as reminders [22]. There are also other reasons why we manage information: we create a set of digital photos whose only purpose is randomly showing in a digital frame, save requests from IM contacts so they are able to start communication (even if we know we will not need that information again), hang information on a wall for decoration purposes (certificates, diplomas, etc), collecting magazines we know we will not be reading again but would like to have a collection completed, etc. In a sense most of this exposed information is placed for later retrieval as well but its main purpose is something else. If we summarize the above, PIM can be described as management (handling, storing, classifying, organizing, archiving) of personal information by a person for various purposes (later retrieving, reminding, collecting, decorating etc.) to support needs and tasks.

In the next subsection two PIM frameworks are described.

## 2.2 PIM activities

To understand how PIM is performed, Barreau tried to dismember it and so divided it in 5 sub-activities [4]:

- Acquisition: deciding which information will be included in information space, defining, labelling and grouping information.
- Organization and Storage: classifying, naming, grouping and placing information for later retrieval.
- Maintenance: updating out-of-date information, backing up information, moving or deleting information from information space.
- Retrieval: process of finding information for reuse and
- Output: visualizing the information space based on users' needs and objectives.

This classification of PIM activities was a basis for Boardman's classification [10, p. 17]. He argued that updating information content cannot be a part of PIM, as it deals with content of information items. He also argued that visualizing is done by computers (not users) and that visualization is present in all sub-activities. He describes four PIM sub-activities as:

- Acquisition: naming and/or (deciding of a) placement in information space.

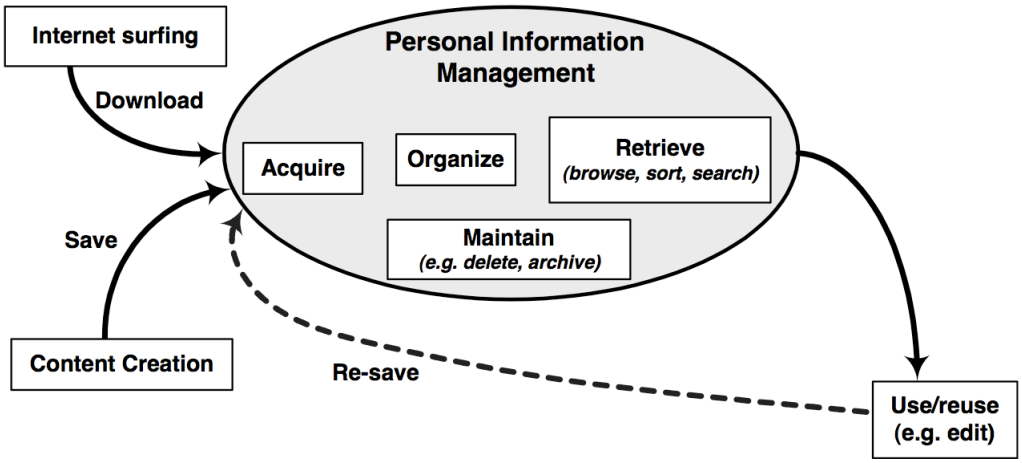


Figure 1: PIM activities and activities outside PIM by Boardman

- Organization: placing information items, renaming, moving and creating new folders.
- Maintenance: backing up and deleting information from information space.
- Retrieval: browsing, sorting and searching for information.

Boardman's view on what activities are a part of PIM and what activities are not (creating, (re)using, downloading and (re)saving information) can be seen in Figure 1 [10, p. 17].

Jones and Teevan grouped all PIM activities in three main groups that support our needs in correlation to information [17, p. 13]:

- Keeping activities: decisions focused on a single information item about the future needs and future availability.
- (Re)finding activities: driven by our needs for information in PSI.
- Meta-level activities: maintenance (composition and preservation) and organization (selection and implementation of a scheme) of the PIC within PSI, managing privacy, evaluating PSI, making sense of information and information distribution.

The first group of activities is focused on the flow from information to our needs and the second group on the flow from our needs to information. All other activities support both flows as can be seen in Figure 2.

These two PIM frameworks set up the basis for conceptual description of PIM. More structured description of how information enters in our digital PSI on personal computers is described in the next section.

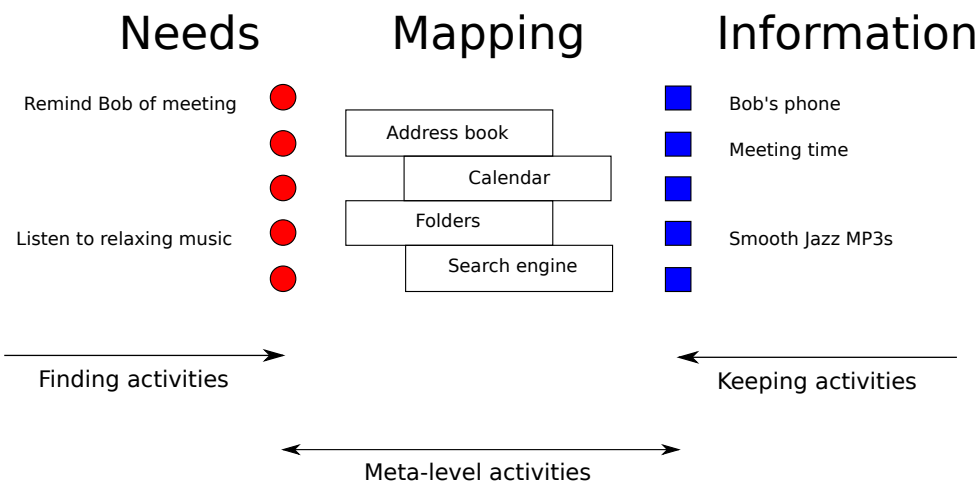


Figure 2: PIM activities viewed as an effort to establish, use, and maintain a mapping between needs and information

# 3 Information importance and mental information linking - hypotheses

In previous section (2.2) we looked at Boardman's, Jones' and Teevan's view of PIM and its activities. In Figure 1 Boarman includes internet browsing and content creation as information sources to the acquisition. On the other hand, Jones and Teevan set our needs as a trigger which helps us decide which information to include in PSI. This step will be described in more detail in the next subsection to fully understand how, when and why users acquire and organize their information and because the type of acquisition directly affects all other PIM activities which we will present in 3.2.

## 3.1 Information acquisition

We will first divide acquisition of information in our digital information space as type of acquisition affects PIM activities and subjective importance, it might trigger new tasks and it also affects acquisition of new information:

- Manually acquired information.
- Semi-automatically acquired information and
- Automatically acquired information.

We have a total control over *manually acquired* information which can be (re)named and/or placed in information environment based on our decisions at the time we decide. On the other hand, *automatically acquired* information piles up in a predefined place while we do not have a control over its acquisition. While *semi-automatically acquired* information still needs our action to be acquired but some actions (like naming and placing, time of acquisition) can be done without our intervention (e.g automatically by an application).

There is another way of looking at information acquisition.

- Intentionally acquired
- Unintentionally acquired

The first group includes information we intentionally acquire. Our actions to get that information to our PSI based on our needs are intentional. It can be noted that intentionally acquired information includes manually and semi-automatically acquired information. While automatically acquired information is always unintentionally acquired.

Automatically acquired information represents a great proportion of information that enters in one person's PSI. Examples are a flow of emails, RSS feeds and calendar inputs from others. This *information is directed to a person* by others - humans or machines (see definition of personal information in Subsection 2.1).

Automatic acquisition is not based on our needs. The difference between intentionally (manually and semi-automatically) and unintentionally (automatically) acquired information is the time when a need for it emerges. With intentionally acquired information our needs emerge before acquiring it – the need for information triggers its acquisition. With unintentionally acquired information needs usually emerge after information is already acquired in our PSI. While scanning it we can decide if it will remain in our PSI or not, based on our needs. In both cases our needs can be clear and so the name and the place for newly acquired information or they might not be clear and information can be left in the predefined place (inbox, temporary folder, dump area) for possible future use or be even deleted [17, p. 38].

Looking at Jones and Teevan's activities, there is a difference when keeping activities evoke. With intentionally acquired information, keeping activities evoke before information enters in PSI while with unintentionally acquired information, keeping activities evoke when information is already in PSI. In both cases information gets managed by meta-level activities (organized and maintained) to support information-need relation (see Jones' and Teevan's PIM classification in Subsection 2.2). After managing (automatically acquired) information directed to a person by meta-level activities this information might (if not deleted) become *the information a person keeps, directly or indirectly for personal use* (see Subsection 2.1). Meta-level activities are thus also driven by our needs to simplify keeping (making easier to file new items) and finding (making easier to find acquired items) activities.

The information we acquire either way can be divided into three main groups:

- Created information
  - Manually acquired: files we create and store in PSI, calendar notes we create in a certain place(s) in calendar, etc.
  - Semi-automatically acquired: emails we write to others that are automatically placed in a sent folder after sending, emails we write to ourself (and know they will be placed in inbox), URLs we visit knowing they will be placed in a web browser's history, etc.
- Received information
  - Manually acquired: information received on removable media (CDs, DVDs, USB sticks) and manually placed in our PSI, etc.
  - Semi-automatically acquired: emails we ask other people to send us with information we need (not knowing the time of acquisition), etc.
  - Automatically acquired: received emails (and attachments) by others that we do not expect, started IM chats by others, calendar inputs by others in shared calendars, RSS feeds in a RSS client, etc.

- Found information (outside our PSI)
  - Manually acquired: information such us files we find on the internet and manually place in PSI, etc.
  - Semi-automatically acquired: information such us downloaded files from the internet that are automatically placed in a predefined folder by web browser, files that P2P clients store in a predefined folder, automatically piled web bookmarks on te first level of a hierarchy, etc.

Even if some applications allow us to set up organizational rules (for example placing emails in folders based on the sender), we do not have a control over the time and amount of incoming information; acquisition does not require our action (clicking on a link, asking for information, ...). Actions such us opening an email account, subscribing to RSS feeds, etc. are in our opinion not parts of acquisition even if Jones describes them as keeping activities in a broader sense [17, p. 38]).

There are also other factors that affect acquisition and organization. Boardman suggests that (1) personal factors such us tidiness affect organization activities and (2) that files and bookmarks (manually and semi-automatically acquired information) organizational structures grow incrementally and are easier to organize than emails (automatically acquired information) [10, p. 111]. This again suggests that users are more likely to manage (name, place, back up ...) manually acquired items at the time of acquisition and manage (if they do so) automatic flow after acquisition.

Based on the above separation of acquired information we will reflect on other PIM activities (organization, maintenance and retrieval) and reconsider the PIM framework in the next subsection.

## 3.2 Reconsidering PIM activities

Decisions like ”store this document” and ”name it” make us think of present organizational schema of PICs and PSI. How is an information item fitting in the present schema? Is present organization supporting our needs (tasks) for storing it. Should we create a new folder, group it with some items, reorganize a part of a schema or just dump information in a temporary folder? All these questions (and possibly others) arise while placing a new information item in organizational schema and actions taken change the state of (organizational) structure of PSI. These activities are, in our opinion, part of organization. As such the border between acquisition and organization is not very clear. While acquiring we have to decide on a name and place (filing or piling) of new information items, we might create a new folder or perform more radical changes to organizational structure. Also border between organization and other (PIM) activities is not very clear. For example: while retrieving a document we realize that it needs to be grouped with some files in a new folder to support our needs and thus tasks. Performing PIM activities can result in changes of the organizational structure of PSI and as such these activities include some organization as well.

PIM activities rarely take place one after another. More often PIM activities overlap with each other and other activities. Besides that performing some PIM activities can not be completed without involving organization (like manual acquisition), PIM activities can overlap because our needs overlap as well. While performing one activity a new activity can take place before the first activity ends as seen in Figure 3. For example while organizing emails (a present need), a need emerges such us ”I have to remind Bob of a meeting” (a pre-existing need we forgot about) or while browsing for a document in a file system (a present need) we come across MP3 files that trigger our need to listen to music (a new need). So needs for information can emerge during all PIM activities as well as needs for information are driving us to perform PIM activities. As described, these

overlapping PIM activities might not even be related to the management of the same information item(s) and run concurrently with other activities we perform in PSI (reading, editing, watching, listening, chatting ...).

Figure 3: On the right are four overlaping PIM activities - one activity might not be finished yet when new activity starts. On the left PIM activities happen one after another. Other non-PIM activities can happen concurrently and overlap with PIM activities which is not shown in the figure.

Note that our view of organization activities differs from Jones' and Teevan's. In their opinion organization activities take place while we think about the whole organizational structure of PICs and happen once so often [17, p. 39]. However, we think that users change organizational structure (even if only a small part) also during other activities and that many such small changes incrementally (while thinking and rethinking about small portions of the existing schema) change structure and organization of PICs and PSI.

While organization deals with information items related to present ongoing tasks, maintenance is performed to support future (maybe unknown) needs and tasks. Maintaining means to manage the organizational schema in such a way that information related to past tasks, for which it is assessed that it will not be needed any more, is moved out of the way and the organizational schema is organized in a way to support future tasks while it still has to support ongoing tasks. Maintenance assures that organizational schema and existent information assist our needs and tasks in the longer run. Even here the border between maintenance and organization is not clear. During the maintenance users perform some organization activities (moving, deleting, creating folders and items, grouping, renaming, etc) and other activities such as archiving (moving out of sight either by moving information on removable media or hiding it deeper in a hierarchy), synchronizing and backing up.

Based on all said, we therefore propose few modifications to Figure 1 as summarized below and as seen in Figure 4.

- Three sources of acquired information enter PSI: created, received and found.
- There is no clear border between organization and other activities (manual acquisition, retrieval, maintenance) and as such these activities intersect. But organization activities can also happen alone (e.g. we create a folder structure in advance (prospective folder creation[18]) as we know we will need it later on).
- Naming and placing information items are two sub-processes of saving and downloading and as such part of organization (or manual acquisition).
- (Re)using information deals with the content of information and can be directly followed only by organizing (if for example information is moved elsewhere, renamed or other data which affects its position is changed - e.g. date in a calendar or surname in contacts). Otherwise this information is already acquired and editing or viewing its content does not affect its name and position in PSI.
- Several (clear and unclear) needs (that are described above) are driving us to perform activities and emerge while performing activities. Automatic acquisition is not affected by our needs while organizing automatically acquired information is.

- Reuse can follow any of four PIM activities. For example while maintaining file system hierarchy we open a file to see its content and decide about deletion or while organizing we (re)read emails to decide about a place in a hierarchy or we simply retrieve a file to view/edit its content.

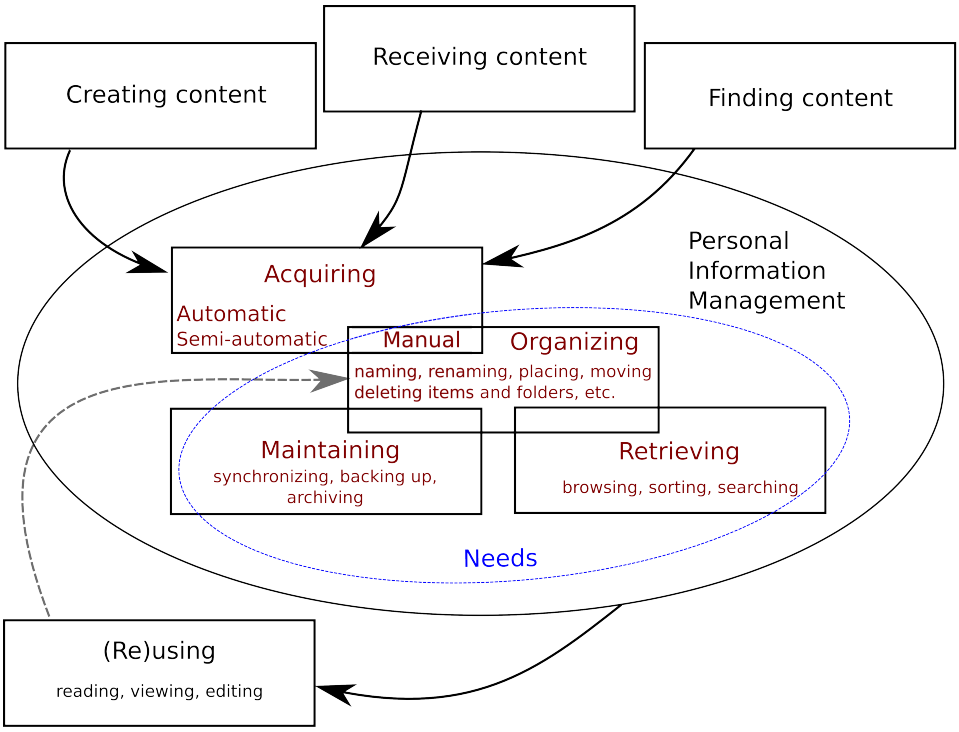


Figure 4: PIM activities revised

Described PIM activities, information items importance and its relation to creation of cross-tool information collections are described next, based on the type of acquisition and all proposed changes to the two PIM frameworks.

## 3.3 Information items importance

It is well known that objects we build or create have greater value over objects we buy or receive [23]. Similar conclusions have been made about PSI as well [10, p. 110][3]. Barreau and Nardi divided information items, managed by participants in their study, in three types based on their lifetime in the forefront of PSI: *ephemeral*, *working* and *archived*. They noted that participants regarded ephemeral and working information as more important than archived information. Although information importance was mentioned by participants and researchers in several studies it was stressed out by Bergman et. al. They mentioned subjective importance of information items as one of the three user-subjective principles (besides subjective classification and subjective context) that should be considered in the design future PIM tools [7].

Boardman reported that users tend to spend more time organizing files and do it more extensively than for email and bookmarks. He suggests that: (1) users spend more time organizing files

because they feel stronger ownership and (2) users think that file re-use is more likely than the re-use of email or bookmarks [10, p. 110]. He grouped information by two key factors [10, p. 202]:

- Information usefulness - *active* (ephemeral, working items), *dormant* (potentially useful), *not useful* and *un-assessed* (unread emails)
- Information ownership - *mine* (self-created items and items with value such us emails we file in a folder) and *not mine* (information on the internet, great proportion of emails in the inbox, etc.)

The adjective *important* is defined as *marked by significant worth* or *valuable in content* by the online Merriam-Webster English dictionary (in this paper importance is understood as subjective importance to its user). We can therefore assume that information items we create are more important than items we find or receive. If we, for example, delete a file we spent 8 hours creating (editing), we loose 8 hours of our work. If we delete a file we spent 8 hours searching for it, it will probably take less time to find it once again (which also depends on a time between finding and re-finding). If we delete a received information item we can ask for it again. There are also exceptions. In our PSI often come information items that expect our immediate response. Such items might have a high importance level at the time of their acquisition but the level can also quickly drop as we respond to them. An example could be an email in which our superior is giving us some urgent work to do. At the moment the importance level is very high but it drops as soon as we finish that urgent work.

Information items might be considered important for longer periods of time not only because of the added value but also because they relate to tasks that last longer. Importance level of such items also diminishes over time when we finish related tasks and new items gain higher importance level as we start working on new tasks.

So far we listed four possible factors that affect importance level:

- usefulness to the related task,
- ownership (who created it and acquisition type),
- added value thought PIM and other activities (invested effort and time to creation, editing, organizing, maintaining, searching, etc.) and
- essentiality and duration of the related task(s).

These factors might or might not be interrelated but they all affect the assessment of information importance. Besides the assumption that created information is more important than received and found, we can also assume that manually acquired information items are more likely to be considered important (and more carefully managed) than semi-automatically and automatically acquired because of the ownership and/or added value. The type of acquisition is thus related to importance assessment. But over time some initially automatically acquired information might get added some value through modifications and ownership level might raise as well. Management of acquired information is so affected also by importance assessment this items might have in the future (clear future needs). Our first hypothesis describes the relation between level of importance of information items, type of acquisition, factors that affect importance level and ongoing tasks defining lifetime of information in the forefront of PSI.

*Users asses their newly acquired information as important and unimportant. Level of importance of an information item in PSI is initially based on acquisition type and assessment of four factors:*

*usefulness, ownership, added value and task(s) essentiality; but can change over time as assessment of these factors change. Level is higher if a combination of factors that affect importance level score high and such information is likely to be managed. Finished the related task(s), importance level drops and information is likely to move out of sight (be archived).*

## 3.4 Information items mental linking and formation of Task Information Collections

Several studies were carried out and solutions were proposed to help fight fragmentation of semantically connected information on personal computers. Despite managing three hierarchies (files, emails, bookmarks) that hold semantically connected information items, it was found out that most folders in different hierarchies do not overlap as users may have different organizational needs for each collection [10, p. 100]. When surveyed about PIM-tools problems users reported that [24]:

- the separation of files, email, and bookmarks is inconvenient since together they form a complete PSI;
- there is no support for short-term notes or remarks, and available solutions are not practical or useful and
- there is no link between various resources.

To complete a task, information from various sources is needed. Such information could be only residing in one tool and managed together as a PIC. Creation of a PIC is based on a performed task but it also depends on actual application used to manage it. For example information items in one hierarchy can be stored and managed as a PIC together in a folder which provides a simple yet intuitive way of storing semantically connected information of one technological format (or more if we have a folder of files). Simple problems arise when an item belongs to several folders. But more often information needed to complete a task is managed by several tools. Information related to one task can include sets of files and folders, web pages, contacts, to-do notes, calendar inputs and emails, and new items can enter a collection as well as old items leave it. Such information can reside in several PICs managed in different applications in different ways and can include also individual information items dispersed in (digital and physical) PSI which are not a part of any task based PIC (e.g. a set of all emails in Inbox). We will call this information Task Information Collection (TIC).

There has been a lot of research about how people perform their tasks. Czerwinski et.al. studied how interruptions are affecting the tasks completion [13]. They let users define their tasks (users' subjective definition of their tasks from their perspective is also our understanding of a task) and found out that users have most problems returning to long lasting tasks while most of the tasks were just short lasting from a few minutes to couple of hours. There has been a lot of suggested solution to bring together information that is task related [25, 5, 6, 19, 15, 8]. Some solutions even bend towards helping users to complete their task with suggesting future possible actions [14, 12]. But none of these has studied how tasks and its information and information importance evolve over time and if all information related to a task needs to be noticeable and remembered (computerized) all the time during the task and after its competition, or if managing separate PICs helps keeping the sense of control over small manageable chunks of information space.

With present tools, managing TICs is left to users. PIM applications do not support managing items dispersed over several folders, hierarchies (applications) and devices. Users create mental links

between information items of a TIC that are placed in the same hierarchy but in different folders and across hierarchies and different folders in them. TIC includes also mental links to information items from a physical world, other digital devices and information pieces that are acquired knowledge in users' heads (do not exist in any form in physical and digital worlds). Mental links between items in one TIC and across TICs form a complex interrelated alive environment of information items. The problem is even more profound as TICs can also overlap. An information item(s) or even folders can be a part of several TICs at the same time (a simple examples can be a contact with whom we work on several projects or a folder of files used in several tasks like collection of papers we use for referencing in several papers).

Memorizing all mental links might lead to a mental workload. We think that with a vast amount of information items that can possibly be mentally linked, users actually mentally link only information that seems important to them at a certain time and these links constantly change as new items gain higher importance level and older items lose it. Lets consider a task of writing a paper. We could for example place this paper in a folder and have a subfolder of photos associated with a paper. This links are managed by the file manager and are supported by a parent folder. This paper is also mentally linked to several email messages (and contacts in address book), some papers placed on the web (as references) and a conference web site. At the moment of writing and submitting a paper, all mental links are clear, but after a year or two most of these links vanish and only few links might be remembered if at all (maybe links to items in the same folder as our paper). So TICs are forgotten as their related tasks are completed and new TICs emerge as new tasks start and these TICs that are present in a persons memory is ever-changing until it is forgotten. Note that while PICs are more static as they are somehow dependent on a tool, TICs are more pliable as they depend on our efforts to sustain them in our minds.

To surpass these weaknesses users have to somehow manage cross-tool management. The second hypothesis defines relation between suggested TIC creation (mental links) and importance:

*Information items that form TIC across tools are mentally linked, managed by several tools and can reside in several PICs and can also overlap with other TICs. The burden of linking information across tools and remembering these links is left to users. Thus users mostly remember links between information items rated as important at a certain time and do not bother with links to and from unimportant information. While importance level drops (after a finished task), links are forgotten as well.*

To prove (or disprove) our hypothesis we built an online questionnaire that allowed participants to enter data about information they regarded as important. Design of the questionnaire and the methodology are presented below, followed by the study description and results.

# 4 Crosstool study of important information

In this section we will present methodology and how the study was designed (4.1), description of the process of the study (4.2) and results by information type (4.3 ) followed by the discussion of these results at the end (4.4).

## 4.1 Methodology and study design

In our preliminary study we wanted to gather these data:

- Which items are mostly regarded as important: created, found or received?
- How much time users spend on creating (editing), finding or reading items regarded as important?
- How many (important) items users consider linked to other items and do linked items reside close to each other (in one hierarchy or similar places across hierarchies)?
- Do users link more created, received or found items and do users link more files, emails or web pages (bookmarks)?
- How long do important items bear this label?

While deciding on methodology of how to gather this data, our first concern was that users would not like to have software installed on their computers to log what information items they are using. So we decided to build an online questionnaire which enabled participants to enter data about information items they felt comfortable to disclose. This might be the main weakness of the study as participants were left on their own to enter data which could not be proved. But we wanted the questionnaire to be anonymous and participants were instructed that they did not have to disclose real names or paths of information items. Another weakness might also be that we interviewed participants over emails and not in person. The reason was again that we wanted the questionnaire to be anonymous to see if and what information items would participants disclose in such questionnaire.

The questionnaire was constructed of three parts: login form, calendar and questionnaire itself. Registration was anonymous but participants could enter their (temporary) email addresses and names for communication purposes (they could still send anonymous comments over an online form). The obligatory data were age, education, job position (if working), dominant file manager(s), email client(s) and a web browser(s). Before logging in, participants were informed about the purpose of the study, the privacy of their data and shown an introduction on how to use the questionnaire.

After logging in, participants were showed a calendar. On each day in calendar were three URLs, each leading to a questionnaire. One questionnaire was for files, one for emails and one for web pages (bookmarks). Participants were asked to enter as many information items as they considered important each day. If a participant filled in information on a certain day and technological format the red dot switched to the green one in front of that technological format (URL) on that day. Data for already filled information items could be reused on other days to simplify and speed up insertion. The calendar and the questionnaire can be seen in Figure 5.

When entering information about an item, participants were asked if this item is related to other items (files, email, web pages) not placed in the same folder. Participants could choose between already entered important information items or insert new not (yet) important items (which could also be reused and gain importance label). Answering this question, links between items were recorded. A link from a file to an email was recorded, for example, when participant entering information about a file related that file to an email. If, while entering the information for that exact email, this email was also chosen to be related to the same file, the link was created in both directions (from a file to an email and from an email to a file). We will refer to these links that participants created in a questionnaire (which were links they mentally created in their heads), when talking about links between different items in Subsection 4.3,.

Before starting our study about importance and cross-tool TIC creation, we showed the questionnaire to 3 researchers in the Computing department of the University of Primorska who were asked to use it for a week. After a week we changed the interface and few questions based on their comments, shortened the questionnaire which resulted in less collected data as we wanted to collect

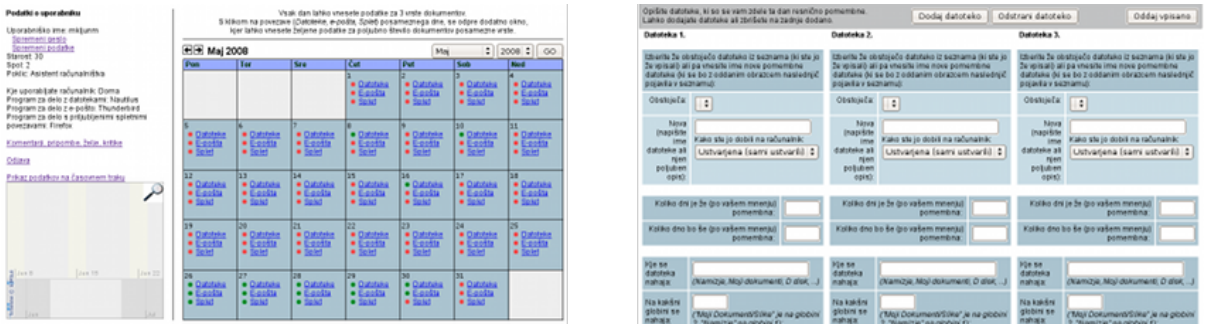

Figure 5: Left - calendar view. Right -a popup questionnaire window as a result of clicking on a link in the calendar

in the first place to prove all our hypotheses. But nevertheless these changes made questionnaire more easy to fill in.

## 4.2 Study proces

Computer science students from the University of Primorska and University of Ljubljana were invited to participate in a study. All obligatory questions were answered by 25 students that registered in the questionnaire. The average age was 22 (minimum 19 and maximum 33). Four were female. One of the participants used a PC only for school and two only for work. Others used it for various purposes.

Windows Operating System (OS) was used by 19 participants, 4 participants used Linux OS and 1 Mac OS X. One participant did not reveal this information. Linux has a higher percentage rate compared to the world usage (Net Applications May 2008), which is probably because participants were all computer science students. The comparison can be seen in Figure 6.

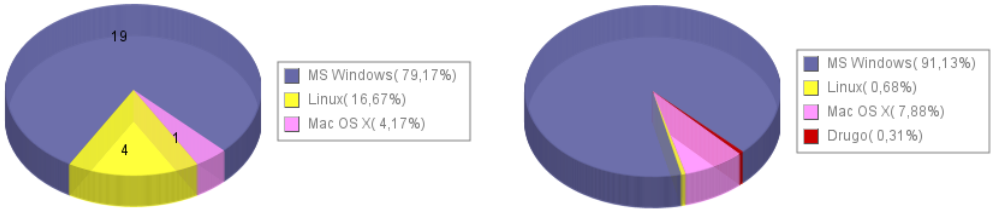


Figure 6: Left - OS usage among participants. Right - OS usage in the world (Net Application May 2008)

Mozilla Firefox was a primary web browser for 20 participants. Only 3 participants used Internet Explorer, 1 used Opera and 1 Safari. The picture differs from world's usage and the comparison can be seen in Figure 7.

Gmail was a leading email client used by 10 participants, followed by Mozilla Thunderbird (8), Outlook Express (4), Yahoo (1), Incredimail (1) and Mailplan (1). It is very interesting that more than half of the participants used a web based mail client which can be attributed to students mobility. The results are shown in Figure 8.

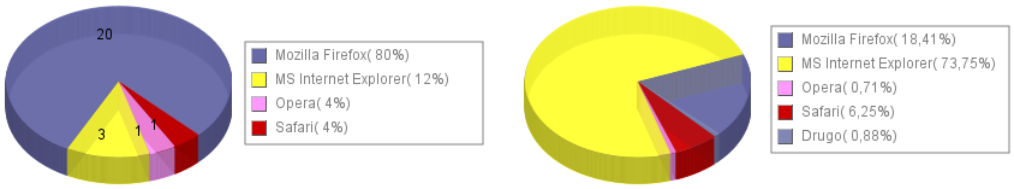


Figure 7: Left - Web browser usage among participants. Right - Web browser usage in the world (Net Application Maj 2008)

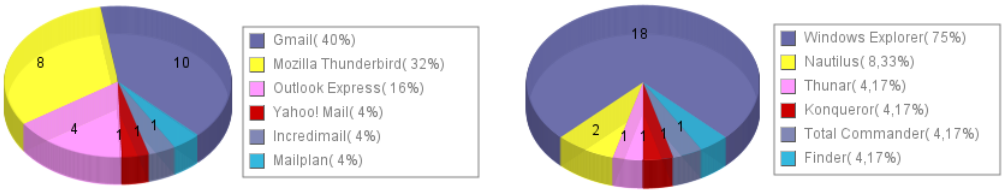


Figure 8: Left - Mail client usage among participants. Right - File manager usage among participants

Most of the users used Windows Explorer as their file manager (18), followed by Nautilus for Gnome Desktop Environment (DE) (2), Konqueror for KDE (1), Thunar for XFCE DE (1), Total Commander for Windows (1) and Finder for Mac OS X (1). One participant did not disclose this information. Results are shown in Figure 8.

The main problems of our study were its longevity and qualitative nature of almost half of the questions in a questionnaire. Even if we tried to simplify it as much as possible, so the time to complete it in one day dropped from 15 minutes to 5 minutes, only 7 participants entered data almost every day for a month. We could still draw some interesting conclusions and find similarities with other studies which are discussed below.

## 4.3 Study results by information type

Below are presented results by three information types: files, emails, web pages. Collective results are presented in the next subsection 4.4.

### Files

All 7 participants entered 39 files (maximum 10, minimum 4). Of these 35 were regarded as important. Other four were entered as mentally linked information to important emails and web pages and were not regarded as important. Participants created 24 files (68.6%), five were found (14.4%), five received (14.4%) and one that did not have this data entered. This is really important as two thirds of the files regarded as important were created which partly confirms our thoughts. Another considerable fact is that it took on average 55 minutes per day to create these files (16 minutes for reading and 1 minute for finding) so the effort was noticeable (see Table 1 about spent time on items that were important more than once). Finding activity did not cause major problems mostly because items were located on the first or second level in a hierarchy (see Table 1) and their lifespan of being important was much shorter than we expected.

Of 24 created files one was regarded as important 15 days (in a range of 28 days), one 10 days

(in a range of 12), one 3 (in a range of 9) and two files 2 days (one in a range of 4 and the other in a range of 3 days). Only one received item was regarded as important twice. The rest of 29 files were regarded as important only one day which we did not expect. We thought that important information will be regarded as such for longer periods of time. Two of the participants responded that they often create short documents which are important only that day.

Table 1: Information about 6 files that were important more than one day: number of days a file was regarded as important, depth level in the hierarchy, type of acquisition (R - received, C - created) and time spent on an item

| No. of days regarded as important | 2 | 2 | 2 | 3 | 10 | 15 |
|---|---|---|---|---|---|---|
| Hierarchy level | 1 | 1 | 1 | 2 | 2 | 4 |
| Acquisition type | R | C | C | C | C | C |
| Reading time in minutes | 0 | 15 | 180 | 0 | 120 | 225 |
| Writing time in minutes | 120 | 40 | 420 | 0 | 320 | 645 |

Nine files were stored on the first level in the hierarchy, eight on the second (10 of these 17 files on desktops and seven in home folders), two on third, seven on fourth, four on fifth and two on sixth. The data was not entered for three files. A connection between number of days a file was regarded as important, depth level in a hierarchy and type of acquisition are presented in Table 1. It can be noted that more than 83% of documents regarded as important more than one day were created. It can be also noted that hierarchy level rises with number of days an item was regarded as important. But several documents that were important only for a day were filed on 3rd, 4th, 5th and 6th level of the hierarchies.

The first two levels were described by participants as easily accessed, spatially located (desktop) and easy to find. Deeper hierarchy levels were used for organizational purposes: *"logically divided by folders"*, *"project based organization"*, *"company expects me to organize this way"*, *"I am used to my organization"* and *"files grouped by domain"*. A lot of folder names revealed a project based organization: *desktop/paper*, *desktop/faculty*, *my documents/faculty/systems2/paper*, *home folder/documents/projects/project name*.

Most of created files (21) were related to study or work. Only three created items were related to hobby. Received (5) and found (5) files were either hobby related or leisure related. In percentage this means that 60% of important documents were related to work or study of which almost 86% (18 documents) were created. This also explains the project based names of folders.

Participants mentally linked 16 files with other information items: nine with email (six linked in both directions) of which five also with web pages (two linked in both directions), one with a web page and six with other files (one linked in both directions). These six files linked with other files were not linked to other information items of different technological format - two of them were linked to the same file and not linked in between, two linked files were stored on the desktop, three linked files were stored in the same folder and two were stored in different folders (desktop and C: drive). Even more interesting is that participants mentally linked files on a spatial surface (desktop) even if it represents a folder and also files in different folders (not subfolder). This confirmed our prediction that files from different (and not sub) folders in one hierarchy can be semantically connected.

## Emails

Participants entered 35 emails (maximum 12 and minimum 2) of which 31 were regarded as impor-

tant and three mentally linked to other information items. Only nine out of 31 emails were created, for four emails this data was not entered. All other emails (18) were received which means that participants received and regarded as important 58% of all entered emails while 42% of emails regarded as important were sent.

Participants spent in average 3:33 minutes reading each of received emails (18) and 11 minutes for writing emails (9) which is a lot less than it took them to write files. So the spent effort was not so significant. None of the entered emails was described as found. It is interesting that participants entered a writing time with received emails (which took them in average 1:30 minutes). Participants described it as answering to received emails. So they somehow considered a received email and the answer to it as the same item which contradicts with the information item description (see Subsection 2.1). Another interesting finding was that it took on average 3:30 minutes to read created (written) emails which was described as reading emails before sending them out.

Only four emails were regarded as important twice. Two in the range of two days (received), one in a range three days (created) and one in a range of five days (created). These two created emails (by same participant) were created and regarded as important but finished and sent a few days later. All others were regarded as important only once which was expected as we predicted that importance level of received items drops quicker than the level of created items (which rises with the time spent on them as well, which can be seen in Table 1). There is no significant relation between the time an email is important and type of acquisition. But we can see that emails were important less days than files.

Participants were asked to estimate how many days were entered emails already important and how many days are expected to be important in the future. Estimation on past importance was entered for 16 emails of which one email was estimated to be important 1 day, three 2 days, two 4 days, one 5 days, two 6 days and others estimated as important but without days given. This can mean that users receive important emails but deal with them few days later. What was unexpected is participants' estimates on how long did they think the entered emails will be important in the future. Two emails were estimated to be important 30 days, one 14, one 10, two 5, one 4, two 2 and 14 one day while in reality only four emails were once again regarded as important. This can either mean that emails contained information that was relevant for estimated periods or participants had problems estimating time or that they simply did not understand the question. Time estimation should be studied in details.

Table 2: Information about 4 emails that were important more than one day: number of days an email was regarded as important, depth level in the hierarchy, type of acquisition (R - received, C - created) and time spent on an item.

| No. of days regarded as important | 2 | 2 | 2 | 2 |
|---|---|---|---|---|
| Hierarchy level | 1 | 3 | 1 | 1 |
| Acquisition type | R | R | C | C |
| Reading time in minutes | 10 | 2 | 7 | 7 |
| Writing time in minutes | 0 | 3 | 12 | 45 |

There is also no significant relation between the time emails were regarded as important and level in a hierarchy (see Table 2). Two emails that were sent were automatically stored in outbox. Only one participant was filing emails. Two received emails (one of which important two days and one important one day) were filled on a third level and three (all important one day) on a second level. These emails were also linked to files, but names and hierarchy levels of folders, in which linked files and emails were stored, did not coincide. All other emails of other participants were piled in inboxes or sent folders. One of the reasons might be the high usage of Gmail web client which

stimulates users to pile up their emails and tag instead of file them. Part of the reason can also be found in answers from participants to a question why are items stored in their folders: *fast access*, *visibility*, *everything is (stored) together*, *undone work is marked as unread*.

Relation to study, work, hobby, leisure and other was more divers: four emails were related to a hobby (two received and two sent), six to work (one received, three sent and two without this data given), four to study (three received and one sent), four to leisure (three received and one sent), 11 related to other areas (eight received and three sent) and four unmarked.

Fourteen emails regarded as important were mentally linked with other information items which is approximately 40%. Six emails were linked to files (five linked in both directions) and one of them also to a web page, six with web pages (one linked in both directions) and two with other emails (one of which was important for two days).

### Web pages

Participants entered 39 different web pages of which 31 were regarded as important and 8 as mentally linked to other information items. The question about acquisition source was probably not well structured. Participants were asked if they created a web page (wrote a blog, posted to a forum, edited a wiki site), found it (through a web search engine, other sites, bookmarks) or received it (a link in an email). The same web site was often marked as created, found and received on different days. So the information entered is not very useful to identify the acquisition source. Even if we did not include these part of results into the account if might just be that users one day open the same web page from bookmarks which they ”find” (e.g. bookmarks toolbar in Firefox), one day they post a comment or write something in a forum and one day get a link to the same web page from a friend/coworker.

Out of 31 web pages regarded as important, 27 were bookmarked. One on the second level in the hierarchy and 26 piled up on the first level. Similar to emails participants were not filing bookmarks. Two of the pages regarded as important were set as home pages and for two pages participant wrote the URL address directly in the address bar (*”these two web addresses are to easy to remember to be entered as bookmarks”*).

More interesting is the importance time. Most of the important web pages were regarded as such all month (on average 12 days). This means that participants returned to web pages almost every second day. Two participants estimated that their four pages regarded as important will preserve this label for the next 300 or 999 days which could mean forever. All other estimations were left blank even if participants regarded same pages as important almost every other day. Probably they did not bother entering this information.

One of the web pages regarded as important was a search engine and interestingly none of them were social networking sites. Five of them were connected to study, five to hobby, three to leisure, four to work and 14 to other activities. Participants spent approximately 16 minutes reading web pages (maximum 60 minutes and minimum one minute) per day.

Five web pages were mentally linked to files (two in both directions while three were not regarded as important), eight to emails (one in both directions and three web pages to the same email) and two web pages to the same web page.

## 4.4 Study discusion

Significant findings of this study are summarized below, followed by discussed grouped results of all three hierarchies together.

- More than half of information regarded as important was created;
- Level of importance dropped quickly for most information items;
- Mental links to information items not regarded as important are also maintained;
- More than half of all entered information was linked;
- Application specifics affect PIM practices;
- Mental links between information items form TICs across tools;
- PIM practices (piling, filing) do not prevent (or greatly affect) TICs creation across tools; and
- The evidence that TICs overlap.

Almost 70% of all files regarded as important were created, while almost 60% of important emails were received. If we look at both collections together 32 items were created, 23 received and only five found. So out of all information items regarded as important 53% were created. Relation between number of items and acquisition sources can be seen on Figure 9. It is very interesting to notice how importance of created and received files and emails is almost turned around.

Unexpected finding was that of 32 created items only 22% were regarded as important more than one day. Level of importance of all information items dropped really quickly which we did not anticipate. Only five created files (2, 2, 3, 10 and 14 days) and two emails (2 and 2 days) were regarded as important more than once. This can imply that participants (students) were not involved in many longterm projects.

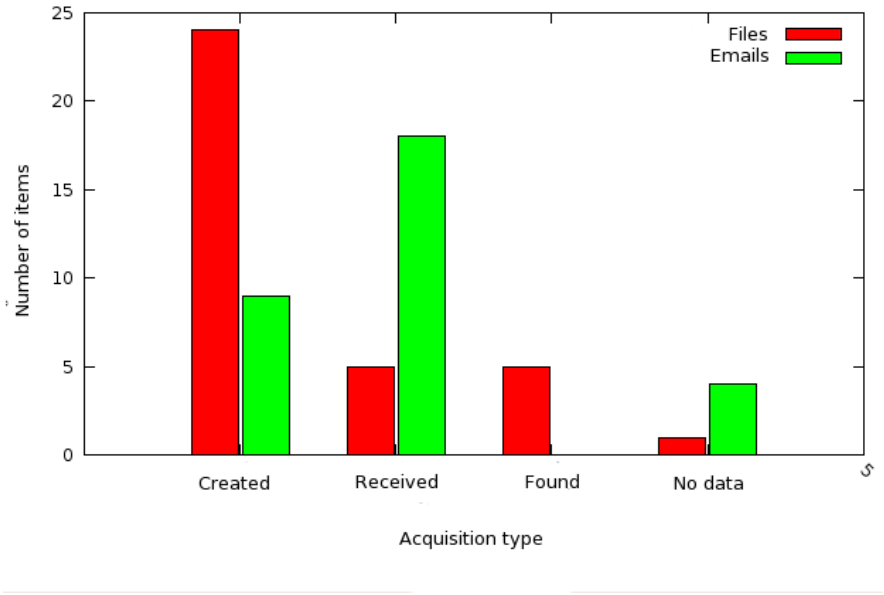


Figure 9: Number of items based on acquisition type

Participants mentally linked information items from three main hierarchies (see Subsection 4.1). Summary of mental links between items can be seen in Table 3. In the first three columns are numbers of links between different types of information items. It must be mentioned that the number of links from one type to another (such us files to emails) can differ from number of links

in the opposite direction (such us emails to files) as some links were pointed in both directions and some only in one. In the forth column are numbers of all entered items and in parenthesis items regarded as important. The fifth column has number and percentage of all linked items. Number of linked items can also be different from the total sum of numbers in first three columns, as some items are linked to more than one item. In last column are numbers and percentage of linked items that were regarded as important.

Even if we hypothesized that mental links to unimportant information are not maintained as a consequence of information overflow, participants still linked approximately 10% of items regarded as unimportant. This disproves a part of our hypothesis and how quickly are links to unimportant information forgotten compared to links to important information should be studied in details.

Table 3: Summary of links between items

| | Files | Emails | Web pages | No. of items (important) | Linked (%) | Linked and import. (%) |
|---|---|---|---|---|---|---|
| Files | 6 (15,38%) | 9 (23,1%) | 6 (15,38%) | 39 (35) | 21 (53,84%) | 16 (41,02%) |
| Emails | 6 (17,14%) | 2 (5,71%) | 7 (20,0%) | 35 (31) | 20 (57,14%) | 14 (40.0%) |
| Web pages | 5 (12,82%) | 8 (20,51%) | 2 (5,12%) | 39 (31) | 19 (48,71%) | 9 (25,71%) |

It can be noted that 53% of all entered information items were mentally linked. Participants mostly linked files to emails and emails to web pages. The results are similar to those of Boardman's study where participants doubled mostly folders from a file hierarchy to email hierarchy [10, p. 159]. But in our study hierarchies could not be compared since only one participant filed emails and none of them filed web bookmarks. Another interesting observation is the difference between the number of all linked web pages (19) and the number of linked web pages that were regarded as important (9). Participants linked several web pages to other information items but did not consider them important. Half of the web pages regarded as important were left unlinked.

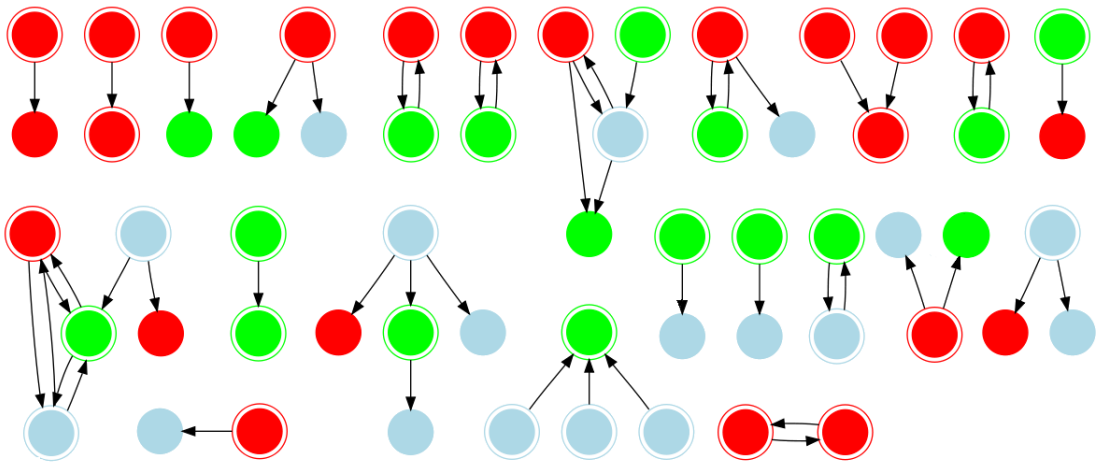

Figure 10: Links between items of all 7 participants. Red circles represent files, green emails and blue web pages. Items with double border were regarded as important. Arrows show the direction of links

All mental links between information items of all seven participants can be seen in Figure 10. Files are represented by red, emails by blue and web pages by green circles. Double edged circles

represent items regarded as important. Arrows show the direction of linkage. Unimportant items have pointed arrows only to them. This is because the questionnaire enabled to link only important information to unimportant (and important) and not the other way around. The purpose of Figure 10 is to show what can be expected when building a support for linkage of simple and complex collection structures. Unfortunately no significant pattern of linkage can be seen.

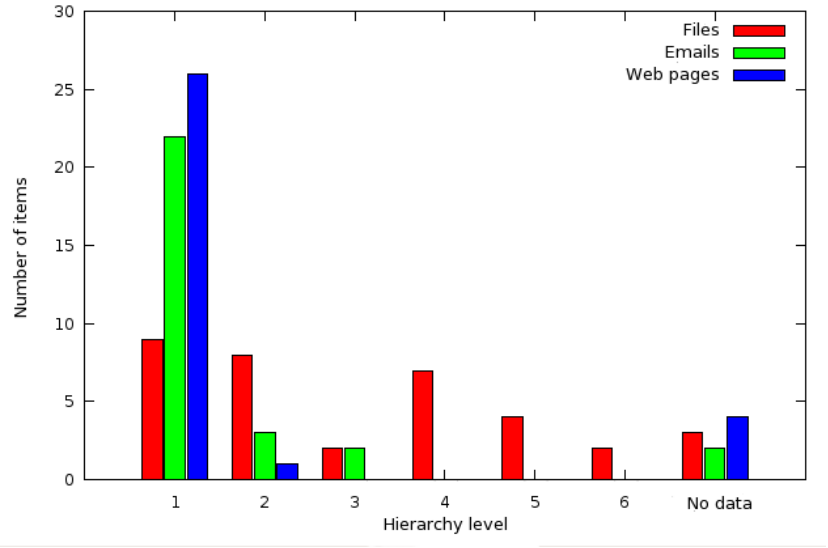


Figure 11: Number of items in hierarchy levels divided by files, emails and web pages

Half of the important files were filed in the first two hierarchy levels (not counting the three items for which this data was not entered). This means either on the desktop or a folder on a desktop or home folder or a folder in it. With the level depth the number of files per folder slowly drops which can be seen on Figure 11. Most of the emails and all bookmarked web pages were stored on the first level as participants mostly did not file them (only two emails out of 31 and one web bookmark out of 31 were filed). This is interesting as these items were regarded as important but still piled up. As Boardman suggested [10, p. 111,202] possible reasons could be personal tidiness, ownership and usefulness (importance factor). But, as we already mentioned, the reason might also be that 6 participants (out of 7) used web email clients that support piling and incorporate a good search engine and bookmarks toolbar in new web browsers. This is an evidence that PIM practices can be also greatly affected by applications specifics. Besides importance factor and application specifics email filing (automatically acquired information) can also be affected because a lack of control over acquisition and amount of received emails (automatically acquired items) [16].

Participants assessed their information as important and unimportant which proves a first part of our first hypothesis. We could also say that who creates information items (mostly files) did affect importance level and that importance of received information dropped quickly. What we could not prove is that information items regarded as important are in the forefront of PSI for longer periods of time (based on the task) an that users move items out of sight when importance level drops. There are some indications of this in a file hierarchy with few files that were important for more than one day, but more studies need to be done.

In Figure 10 can be seen how TICs are created across folders and information types and can not be managed by one tool (application) only. This proves a part of our second hypothesis. It was already mentioned that items residing in the same folder were linked (Desktop), items in different

folders in one hierarchy (not in parent (folder) - child (subfolder) relation) were linked and items from the same folder (Desktop, inbox, sent mails folder) were linked to several TICs. This last is very interesting as items piled in the same folder took parts in different TICs which could mean that management practices do not affect creation of TICs as much (although we still think that management practices help maintaining them). We could then argue that linked items stored in different folders (in one hierarchy and not in parental relation parent-child) prove the overlapping of TICs as each folder can be a TIC on its own. Or that items stored in the same folder linked to different TICs (across hierarchies) also prove TICs overlapping. But more studies need to be done.

# 5 Conclusion

The two main aims of presented study were (1) to find out how users cope with the information overflow with assigning importance level to information items and (2) how users cope with creation of Task Information Collections (a collection of information items relating to a task) across folders and applications with creating mental links based on information subjective importance . We think that subjective Importance of information items is based on usefulness, ownership (acquisition), self added value and present tasks to be accomplished. All criteria can change over time and affect the importance level of information. But initial assessment of importance of newly acquired information is very crucial as it affects naming and/or placing of information items, which also affect other PIM sub-activities and especially retrieval. Unfortunately we did not try to find out if these factors are really the ones that affect importance assessment.

Information items in digital environments on personal computers are acquired manually, semi-automatically and automatically. We have at least some control over the first two (deciding what goes in PSI, items' names and positions), but no control over the automatic flow of information (email messages, RSS feeds, etc.). We claimed that users would care more about manually created information and regarded it as more important. This was true for files, although received emails took a considerable share of important items. We anticipated this, but also thought that importance level of received information would drop quickly. It was surprising that almost all information regarded as important lost its importance after a day or two. Only few items were considered important more than two days. However, more than half of all entered information items that were regarded as important, were created.

Participants mentally linked information from three main hierarchies to form information collections. More than half of entered information items were linked. Collections took various shapes and sizes but the amount of the available data was to small to notice any pattern. Although participants had to enter the data about information items they considered important, they had a possibility to link these items to any information (also information not considered as important). We hypothesized that with information overflow users try to remember only links between information regarded as important, but a tenth of all linked information was not considered important. This indicates that information items from one collection have different importance levels. Another finding was that, besides collections being built of several information items from three hierarchies and from items in the same folder (desktop, inbox, etc.), a collection form items placed in different folders in a file hierarchy was created. This can be seen as a collection overlap as each folder is its own collection.

Participants were (in several studies) grouped by their approach to information management [1, 2, 11, 22]. However in our study only two general groups of management could be observed with information items regarded as important. These two groups were filers and pilers. All 7 participants filed files but only one of them filed few emails. In email and web bookmarks management the piling technique prevailed. Besides importance and type of acquisition (manual vs. automatic),

the specifics of PIM applications also affect PIM strategies as email clients and web browsers used by participants supported piling as the main storage technique. But linkage of items across tools was not related to PIM strategies.

Presented study answered few questions but other questions raised up as well. It is still unknown if users consider their information as important based on the mentioned criteria: ownership, usefulness, added value and related task. There could be other criteria that affect importance assessment but we did not include this in a questionary as it already took considerable time to complete it every day. We are planning to research this in our future studies. It is also unknown how linkage change over time and if users tend to forget links over time when information items are not any more considered important (after the related tasks are completed). We could not answer this because our study lasted for a month, it was anonymous and the number of entered information items was not high enough. However, this could be answered by installing software on personal computers that would automatically gather information about usage of individual items (which we plan to do in the future). What we could prove was that collections are forming across tools but this raised another questions such us if collections do coincide with tasks. There is a considerable research going on about task management across digital devices [28] which could be helpful to understand the dynamics of Task Collections creation and their collapse (with time passing by). Another question was raised about granularity of information. Several attempts to provide a PIM framework describe an information item as a self contained unit of information that could be managed, grouped, given attributes and manipulated. But our study showed that users consider received email and the answer to it as the same item. So the question is, if users link information items or parts of information items (eg. a file linked to a graph in another file or a paragraph on a web page instead to be linked to a whole file or a whole web page). Although a way to link (parts of) information items can be copying and pasting, it was debated that this approach looses richness or completeness as well as source of original information [20].

Our future work will take in consideration a task management in relation to Task Information Collections creation and their dynamics over time. We will also try to answer questions on how importance assessment is related to both tasks and collections and how users understand information items (which apparently is not the same of how technology enables users to create them).

# 6 Acknowledgements

We would like to thank all the participants that spent a significant amount of time entering their personal information in the questionnaire. We would also like to thank dr. Andrej Brodnik and others who helped us with this research. Nevertheless I would like to thank the Research Development Programme team of Lancaster University for helping me improving my writing style and skills.

# References

[1] D. Abrams, R. Baecker, and M. Chignell. Information archiving with bookmarks: personal web space construction and organization. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 41–48, New York, NY, USA, 1998. ACM Press/Addison-Wesley Publishing Co.

[2] O. Bälter. Strategies for organising email messages. In *In Proceedings of HCI'97*, pages 21–38, London, United Kingdom, 1997. Springer.

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

[4] D. K. Barreau. Context as a factor in personal information management systems. *J. Am. Soc. Inf. Sci.*, 46(5):327–339, 1995.

[5] V. Bellotti, N. Ducheneaut, M. Howard, and I. Smith. 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, 2003. ACM.

[6] V. Bellotti and J. Thornton. Managing activities with TV-ACTA: Taskvista and activity-centered task assistant. In *Personal Information Management Workshop, SIGIR*, volume 4. Citeseer, 2006.

[7] O. Bergman, R. Beyth-Marom, and R. Nachmias. The user-subjective approach to personal information management systems design. *Journal of the American Society for Information Science and Technology*, 54(9):872–878, 2003.

[8] O. Bergman, R. Beyth-Marom, and R. Nachmias. The user-subjective approach to personal information management systems design: Evidence and implementations. *J. Am. Soc. Inf. Sci. Technol.*, 59(2):235–246, 2008.

[9] O. Bergman, R. Beyth-Marom, R. Nachmias, N. Gradovitch, and S. Whittaker. Improved Search Engines and Navigation Preference in Personal Information Management. *ACM TRANSACTIONS ON INFORMATION SYSTEMS*, 26(4), SEP 2008. 2nd International Workshop on Personal Information Management, Seattle, WA, AUG 10-11, 2006.

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

[11] R. Boardman and M. A. Sasse. "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, 2004. ACM.

[12] T. Catarci, A. Dix, A. Katifori, G. Lepouras, and A. Poggi. Task-Centred Information Management. *Digital Libraries: Research and Development*, 4877/2007:197–206, November 25 2007.

[13] M. Czerwinski, E. Horvitz, and S. Wilhite. A diary study of task switching and interruptions. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 175–182, New York, NY, USA, 2004. ACM.

[14] A. J. Dix. Tasks = data + action + context: Automated task assistance through data-oriented analysis. In *TAMODIA/HCSE*, pages 1–13, 2008.

[15] P. Dourish, W. K. Edwards, A. LaMarca, and M. Salisbury. Presto: an experimental architecture for fluid interactive document spaces. *ACM Trans. Comput.-Hum. Interact.*, 6(2):133–161, 1999.

[16] D. Fisher, A. J. Brush, E. Gleave, and M. A. Smith. 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, 2006. ACM.

[17] W. Jones and J. Teevan, editors. *Personal Information Management*. University of Washington Press, 2007.

[18] A. Kamaruddin, A. Dix, and D. Martin. Why do you make a folder? Poster, September 2006.

[19] V. Kaptelinin. 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, 2003. ACM.

[20] D. R. Karger and W. Jones. Data unification in personal information management. *Commun. ACM*, 49(1):77–82, 2006.

[21] M. Lansdale. The psychology of personal information management. *Applied Ergonomics*, 19(1):55–66, March 1988.

[22] T. W. Malone. How do people organize their desks?: Implications for the design of office information systems. *ACM Trans. Inf. Syst.*, 1(1):99–112, 1983.

[23] D. A. Norman. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, December 23, 2003.

[24] P. Ravasio, S. G. Schr, and H. Krueger. In pursuit of desktop evolution: User problems and practices with modern desktop systems. *ACM Trans. Comput.-Hum. Interact.*, 11(2):156–180, 2004.

[25] G. Smith, P. Baudisch, G. Robertson, M. Czerwinski, B. Meyers, D. Robbins, and D. Andrews. Groupbar: The taskbar evolved. In *Proc. OzCHI*, volume 3. Citeseer, 2003.

[26] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422, New York, NY, USA, 2004. ACM.

[27] M. Tungare. Understanding the evolution of users' personal information management practices. In *In Proceedings of the Eleventh IFIP TC13 International Conference on Human-Computer Interaction - INTERACT 2007 Doctoral Consortium*, 2007.

[28] M. Tungare and M. A. Pérez-Qui nones. Mental workload in multi-device personal information management. In *PIM Workshop 2009, Vancouver, BC, Canada*, November 7-9 2009.