Downloaded from orbit.dtu.dk on: Mar 16, 2025

DTU Library

NEXUS

A Unified Approach to Personal Information Management in Interactive Systems

Larsen, Jakob Eg

Publication date:
2005

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Larsen, J. E. (2005). *NEXUS: A Unified Approach to Personal Information Management in Interactive Systems*. Technical University of Denmark. CICT Ph.D. series No. 6

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Ph.D. Thesis
CICT Ph.D. Series No. 6

Technical University of Denmark

NEXUS

A Unified Approach to Personal Information Management in Interactive Systems

A thesis submitted in the fulfilment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in
COMPUTER SCIENCE

by

Jakob Eg Larsen

Copyright © Jakob Eg Larsen, 2005

Published by Center for Information and Communication Technologies, Technical University of Denmark, Kongens Lyngby, 2005.

All right reserved. Except for the quotation of short passages for the purpose of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

This book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

Printed by Schultz DocuCenter, Denmark

ISBN 87-90288-21-1

NEXUS

A Unified Approach to Personal Information Management
in Interactive Systems

Copyright © 2004, 2005 by Jakob Eg Larsen
ALL RIGHTS RESERVED

For Stine and Magnus

Abstract

Almost everyone carry out personal information management at one level or another consciously or not. Due to the limitations of human cognitive capabilities different kinds of tools are used to support and facilitate the processes involved in personal information management activities. Research in the area includes observations of personal information management practices in both physical office environments and in use of different computer-based systems. This has lead to an initial understanding of aspects of personal information management and identification of problems involved. Novel models and systems to support personal information management have been proposed. This includes different approaches to the organization of information. Typically different dimensions of memory about information have been utilized, such as, temporal, spatial, and topic-oriented. However, research into the capabilities of these systems in terms of supporting personal information management has received little attention.

This thesis includes a survey of research and studies of the area of personal information management including different approaches to the support of it in interactive systems. Based on these prior results an entity-attribute-collection-action model is described. The model aims at addressing issues and problems in personal information management and provides a foundation on which interactive personal information systems can be built. Existing interactive personal information management systems are characterized by a compartmentalized support of different types of personal information and activities. This approach aims to provide integrated and unified interaction with information in systems based on the model.

A research prototype implementation of an interactive personal information management system has been developed based on the model. The user interface of the research prototype provides unified interaction across information types and information structures in the system. The prototype serves as a proof-of-concept, and allows preliminary experiments to be carried out.

Semi-structured interviews were conducted with eight test subjects to study their personal information management habits and practices. The results generally confirmed findings from prior studies mentioned above, and indicate that problems identified in prior studies are still present even in use of present interactive personal information management systems. Additionally a pilot usability experiment was conducted involving observations of test subjects thinking aloud while performing a set of simulated personal information management tasks using the research prototype. The subjects were capable of using the research prototype for the simulated tasks and understand the underlying model. The experiment provided initial indications that the model provides a framework on which personal information management systems can be built.

Future work includes further development of the research prototype, and real use experiments over longer periods of time where subjects manage their own personal information. Although only simulated tasks were given the conclusion is that the experiments have provided initial indications that the model and the unified approach may provide support for the aspects and cognitive processes involved in personal information management.

Dansk resumé

NEXUS
En ensartet tilgang til håndtering af personlig information i interaktive systemer

Stort set alle mennesker håndterer personlig information på et eller andet niveau bevidst eller ubevidst. Grundet begrænsninger i menneskers kognitive formåen må forskellige værktøjer benyttes for at understøtte og lette de processer der er indgår i håndtering af personlig information (eng: *personal information management*). Forskning indenfor området har bestået i observationer af håndtering af personlig information i fysiske kontormiljøer såvel som ved brug af computer-baserede systemer. Det har ført til en foreløbig forståelse af området, samt identifikation af problemer. Nye modeller og systemer til håndtering af personlig information er blevet foreslået. Dette inkluderer forskellige tilgange til organisering af information. Typisk er forskellige dimensioner af erindring om information blevet benyttet, såsom tidslig, rumlig, og emne-orienteret. Imidlertid har evaluering af disse tilganges understøttelse af håndtering af personlig information og aktiviteter fået mindre opmærksomhed.

Denne afhandling omfatter en oversigt over studier af håndtering af personlig information og forskellige tilgange til understøttelse af dette i interaktive systemer. Baseret på hidtidige resultater beskrives en *entity-attribute-collection-action* model. Denne model har til formål at adressere problemer i håndtering af personlig information og give et fundament hvorpå interaktive systemer til håndtering af personlig information kan konstrueres. Hidtidige interaktive systemer er karakteriseret ved en inddeling og adskillelse i brugergrænsefladen i understøttelse af forskellige typer af personlig information og aktiviteter. Den foreslåede tilgang har derimod til formål at give en integreret og ensartet interaktion med informationer i systemer baseret på modellen.

Et prototype interaktivt system til håndtering af personlig information er implementeret baseret på modellen. Prototypesystemet har en brugergrænseflade med ensartet interaktion med forskellige informationstyper– og strukturer i systemet. Prototypen er ment som en illustration af konceptet og muliggør desuden udførelse af foreløbige eksperimenter.

Semi-strukturerede interviews er blevet foretaget med otte testpersoner for at undersøge deres vaner og metoder til håndtering af personlig information. Undersøgelsen har generelt bekræftet resultater fra ovennævnte tidligere studier. Det indikerer at de problemer der er identificeret hidtil ligeledes er gældende med brug af nuværende interaktive systemer til håndtering af personlig information. Derudover er der foretaget et pilot *usability* eksperiment med observationer af testpersoner der tænkte højt, imens de udførte en række simulerede opgaver i håndtering af personlig information ved brug af prototype systemet. Testpersonerne var i stand til at benytte prototypen til de simulerede opgaver, samt at forstå den underliggende model. Eksperimentet har givet foreløbige indikationer af, at modellen er en brugbar ramme for konstruktion af interaktive systemer til håndtering af personlig information.

Fremtidigt arbejde inkluderer videreudvikling af prototypen, og eksperimenter der involverer brug i virkelige situationer over længere tidsperioder hvor testpersonerne benytter egen personlig information. Trods der kun er benyttet simulerede opgaver er konklusionen, at eksperimentet har givet foreløbige indikationer af, at modellen og den ensartede tilgang understøtter håndtering af personlig information og kan understøtte de aspekter og kognitive processer der er involveret heri.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost I would like to thank my supervisor Michael Rose for his continued support. We have had many interesting discussions throughout the years about the issues of supporting personal information management in interactive systems and many other interesting issues.

A special thanks to librarian Annette Holck Mørk, who has always been very helpful finding articles and books, which were relevant to my research. I owe her a big thank you.

Thanks also goes to the test subjects who took time to participate in the experiments with the research prototype.

Last but not least I owe the biggest thanks to Stine, Magnus, and my family for their continued support.

Preface

In his famous paper "As We May Think" published in 1945 Vannevar Bush described his vision of the "memex" – a device that would supplement human memory and facilitate managing large quantities of information [Bus45]. This paper has often been cited as the first to suggest hypertext, as it suggests, "associative indexing" tying information items together – linking information and creating trails.

It has been inspiring for my work to read the paper from a personal information management perspective. That is, in addition to hypertext Vannevar Bush actually propose an interactive personal information management system – "a mechanized private file and library [...] in which an individual stores all his books, records, and communication". Moreover, Vannevar Bush suggests that the traditional hierarchical indexing of information is unsuitable, as the human mind does not work that way. Association of thoughts is the reason for suggesting the "memex" device, utilizing this property of the human mind.

However, even present interactive systems aiming to facilitate personal information management still mainly rely on traditional information indexing. Thus, in my opinion, the challenge to realize the vision of Vannevar Bush still remains.

It is immensely ingenious, immensely complicated, and extremly effective, but somehow at the same time crude, wasteful, and inelegant, and one feels that there must be a better way of doing things.

– C. Strachey (speaking not of Windows but of the IBM Stretch computer in 1962) [Ras00]

Chapter 1

Introduction

Humans keep and manage personal information as part of both their private and professional lives. Time and effort is spent managing and finding personal information. This includes storing information for later use and retrieving that information again. Both paper-based and electronic-based tools are used to support the human memory and to facilitate the process of keeping and managing personal information.

Over the last decades computer-based systems have been used increasingly for information sharing, exchange, processing, and storage, recently in terms of web-based information systems. A drastic increase in networked resources and network accessible information has been observed. The increasing network connectivity and increased use of networked information appliances have increased the quantity of information available to users. The increased use of the World Wide Web, and other networked systems have contributed considerable to this development. Additionally computer-based systems have been used for asynchronous communication in terms of e-mails, and more recently for synchronous communication in terms of instant messaging systems. The increased use of such systems for information exchange and communication has increased the quantity and variety of information that potentially must be managed by humans using computer-based systems.

This poses a set of challenges for humans to keep and organize the information and it also poses a set of challenges to the design of systems, which may support

and facilitate personal information management and use of this information.

Although personal information management has been studied as a phenomenon for a couple of decades the body of knowledge is characterized by being somewhat fragmented. The results are limited and also the application of results is limited and has not lead to considerable improvements of systems supporting personal information management. Typically personal information management has been studied with focus on specific (isolated) usage, such as, management of documents, e-mail usage, or bookmark management. Moreover the studies are often based on interviews rather than observations of actual behavior. A potential problem is that studying the phenomenon in terms of existing information types and applications may cause a somewhat incomplete understanding of personal information management practices and problems. However, it is recognized that carrying out comprehensive studies of personal information management is both difficult and demands considerable resources. It involves issues such as privacy and problems observing subjects in different locations, situations of use, and using different types of tools. This means preferable studies in non-laboratory settings and over longer periods of time. Such studies demand considerable resources and imply that carrying out studies of personal information management as a whole is a major undertaking. Nevertheless, studies focused on use of specific information and improvements of tools which support management of such information exists. These studies provide valuable insights and understanding of some aspects of personal information management practices and problems. Still further studies are needed in order to acquire better understanding of the fundamental problems of personal information management.

1.1 Personal Information Management

Personal information is information that is kept for our own use. The information may or may not be private. It is considered personal, as it is either a private piece of information or it represents a personal interest in a piece of information that is not private. This means that personal information can be characterized as either being private information or having individual interest. The types of personal information may include addresses, phone numbers, appointments, schedules, reminders, messages, facts, ideas, journals, references, notes, diaries,

et cetera. This means information that is used in activities in everyday life.

As the information is personal humans develop and habituate individual strategies for the management of the information. Such strategies include building piles of documents in an office environment, or creating and maintaining an extensive hierarchical filing structure in an interactive information system.

Personal information is characterized by being organized in an ad hoc manner and therefore being unstructured, semi-structured, informally structured, or formally structured[1]. Studies have shown that individuals may use all these types of organizations, and that the weighting among them is highly individual [Col82] [BN95] [FFG96a]. Some humans can be characterized as "pilers" where most information is unstructured or semi-structured, and others can be characterized as "filers" where most information is formally structured [Mal83].

This means that management of personal information is both highly individual and quite complex. Additionally it includes a set of disparate activities, cognitive processes, and types of information. Some of the activities include task management, scheduling, communication, reminding, and filing information, that is, reading, storing, categorizing, searching, sorting, and discarding information. It involves cognitive processes of recalling, recognizing, and categorization of the information.

Even though a set of disparate information is included it is often interlinked and interrelated information. In a particular context of use a set of disparate information types may be involved. For instance an appointment may include information about people, place, schedule, notes about the appointment, and a reminder about the appointment.

Finally complexity is caused by the fact that the information and information structures may evolve over time. The information and the context of its use may also evolve over the lifetime of the information. This makes the issue of storing and retrieving information more difficult, as information may belong in a different structure than originally intended, or it may be used later in a different context than it was originally intended.

[1] Formally structured refers to a systematic approach to the organization of information

1.2 Human Issues

As the information in question is personal it means that all information items managed and retrieved represents *known* information items. This implies that the individual will remember something about the information and may be able to recollect something from memory about the information. This recollection can include different dimensions of memory about information regardless of the approach taken to organize the information. Even if the information was not organized as such using a particular strategy. For instance the individual may partly remember the content of the information, or the time the information was used or received, the person from who the information originated, or a category under which the information was filed.

As humans remember about information along many different dimensions, these different dimensions of memory may be used in recollecting the information [Lan86b] [Lan88]. The spatial location of an information item is one dimension of the human memory. For instance an individual may recall that an information item was left on the upper-left side of a table, in a particular pile, or on a shelf in the office. He may remember when the information item was last used, which represents a temporal dimension of memory about the information item. Additionally he may partially recollect the content of the information item or the size of the information item by some measure. For instance that is was a document about a particular topic and that it contained about ten pages.

1.3 Supporting and Facilitating PIM

Systems and tools are needed to support and facilitate the management, retrieval, and use of personal information due to human cognitive capabilities and limitations of human memory. Additionally the activities involved in personal information management are complex for the individual due to the quantity and diversity of personal information that may be used in everyday life.

As mentioned humans apply different strategies and a wide variety of systems and tools in order to perform these everyday personal information management tasks [Mal83] [BN95] [FFG96a]. It may include the use of both paper-based and electronic systems. Studies have shown how humans develop individual

strategies for information management in an office environment [Mal83] and using electronic systems as well [BN95].

Information retrieval systems have been researched for decades. Different information retrieval techniques and tools have been studied, that is, the process of doing information discovery from some source [JBD01] and [DCC+03]. Many of these systems focus on finding or discovering information. However, for personal information management the issue is not discovering information as such. Once the information has been discovered it becomes known information, that is, information that has been seen before and used in a particular context. In personal information management the problem is finding and reusing information that has already been seen or handled by the individual. The problem of finding and retrieving such already known personal information has received less attention. Such known information items may be retrieved from a repository in which it was originally discovered or it may be stored and retrieved from a personal information repository.

Therefore the purpose of personal information management systems are to support and facilitate the processes involved in personal information management, that is, allowing easy retrieval and reuse of personal information.

Although personal information management is carried out using both paper-based and electronic tools the focus is on the support of personal information management in interactive systems. That is, a focus on user interfaces and interaction in interactive systems that support personal information management activities and information. Thereby also realizing that some personal information may be handled outside such systems.

Studies of personal information management have identified behaviors and styles. Such studies have also identified different examples of unintended use, as systems have not supported the functionality needed. Current systems are limited in terms of supporting the aspects involved in personal information management. The tools might be used in different ways than intended, as they must handle different situations of use and information than they were originally designed to deal with. For instance the management of different types of personal information can be seen in the use of different software applications.

E-mail systems are designed to handle asynchronous communication. However, studies have shown that such systems are typically used for additional func-

tions, such as task management and information management [Mac88] [WS96]. As the systems were not designed to support this additional functionality, users have to cope with sub optimal support and develop their own strategies and ways of use.

Another example is studies of how humans keep references to information. Recently there have been studies of how humans keep references (bookmarks) to web pages (URL's) [JBD01] [ABC98]. Typically webbrowsers have a hierarchical organized archiving system for such bookmarks. However, studies have shown that this feature is not the most widely used method for archiving and keeping bookmarks to web pages. Instead bookmarks are stored in documents or emails, or in some other way in the context in which the bookmark is used. Apparently such bookmark feature in webbrowsers does not provide the needed support of context and reminding relevant to personal information management.

A more general example is the hierarchical organization of information typically used in different software applications for information management. An obvious example is hierarchical file systems available in modern operating systems. Creating a filing strategy and maintaining the hierarchical structure itself is a difficult and time-consuming task. Each information item may need naming and identification. It may not be easy or relevant to provide a name for an information item. In the physical world humans do not have to provide names for all objects. Filing of information may be ambiguous, that is, it may naturally belong in more than one structure. In file systems files typically reside in exactly one directory (folder), even though multiple organizations may be relevant[2].

This indicates that such hierarchical structures may be too rigid and unsuitable for personal information management and use. Not all personal information needs to be archived or organized into a hierarchical structure. Allowing hierarchical organization of information only may therefore not be suitable, as some types of personal information is more suitably organized into other types of information structures. An information item may more naturally belong in an unstructured collection or into a temporally organized structure. Finally an information item may be used for a short while only and then discarded, why it is unnecessary to provide a name or otherwise organize the information item.

[2]Some implementations of file systems allow symbolic links to files, which then simulates a file residing in multiple structures

Additional flexibility is also needed as the structure of information and its use may change over time and due to the semi-structured nature of some personal information items.

An additional problem is that supporting multiple activities or information types in a software application user interface often means a compartmentalized user interface. This means that different information types are handled by different components in the user interface. This issue often cause user interfaces to include different modes of interaction, which may cause confusion and mistakes [Ras00]. Such separation in the user interface allows little integration of the personal information across disparate information types. As activities related to personal information can be closely related to each other information of different type might belong naturally together, as it is used together in a particular context of use. Thus interactive personal information management systems must provide support for these different activities and the cognitive processes involved. It may be advantageous to provide such support in a unified way, as the information and activities are so closely inter-related.

Different approaches (metaphors) have been tried out stemming from research projects as well as commercial projects. These have tried to address the different subsets of problems created by the changing information environment on the one side, and dealing with the particular problems related to the support of disparate activities and information involved in personal information management on the other side.

Typically these different approaches have focused on different ways of organizing and managing information, such as, hierarchical [JRV+89], temporal [FFG96b] [Rek99], spatial organizations [PF93], attribute oriented approaches [KKB+90] [AKA99] [DEL+00] to mention a few. Each of these approaches provides different levels of support for personal information management and each focus on different aspects of personal information management.

These different examples provide indications of some issues in the design and approaches in terms of personal information management. Together studies of use and the different approaches provide useful information, which can be used in further research of personal information management. It hints that different approaches might provide better alternatives to the support of the personal information management.

This means different approaches to the design of user interfaces and interaction in interactive systems in order to support and facilitate the specific activities, types of information, and cognitive issues involved in personal information management. Thereby enabling better support of the diversities and wide range of activities included in personal information management.

1.4 NEXUS Model

To address these issues alternatives to the underlying system model, user interface, and interaction are considered. We present an approach based on a conceptual model, which serves as the foundation for an interactive personal information management system. The approach includes four basic components:

- Information entity (the lowest granularity information unit representing a piece of information, for instance a fact)
- Information attribute (a name/value pair, that is, metadata that can be associated to and describe information entities). To describe *what* the information is (instead of describing *where* it is)
- Information view (a collection and organization of information entities, that is, a subset of information entities)
- Actions (represent functionality that may be invoked on an information entity or collection of information entities to perform a particular manipulation of those information entities)

This simple attribute-oriented model allows a flexible way of storing information. The information does not need a static predetermined scheme. There are no assumptions about data structures or relationships between information entities.

Rather the model allows any number of attributes to be associated with information entities and any relationship between information entities. Therefore it allows information and information structures to evolve. Also multiple organizations of information entities based on the associated attributes are possible,

which means multiple views of the same information. Others have investigated or used attribute-oriented models for information management [GJSJ91] [HKQ02] [DELS99] [BYJR00] [KKB+90].

The properties of the model are important in order to support and facilitate personal information management. As such the model is used as the foundation for an interactive personal information management system, which aims to support and facilitate the aspects of personal information management mentioned in the previous sections.

1.5 Research Prototype

A research prototype of the model has been implemented. This allows a personal information management application prototype – NEXUS – to be built on top of the model. This application aims to support and facilitate personal information management, i.e. it aims to support and facilitate the particular issues involved in personal information management

The research prototype has been implemented as a traditional desktop application with a graphical user interface. It includes support for a variety of different activities related to personal information management, which means information types such as, e-mail, calendar appointments, contacts, addresses, to-do items, notes, bookmarks, et cetera.

The prototype user interface provides a unified user interface to and interaction with the information in the system. The interaction with information entities and collections of information entities is implemented with a small consistent set of commands, which apply across information types. This means that potential users of the system will only have to learn a fairly small command set, which applies across disparate information types.

As the system is built on top of the model the information types mentioned there is no immediate distinction between these information items. They may be distinguished in terms of the different attributes associated with the information item. As such an e-mail is only considered different from a calendar item in terms of the different attributes associated with the two information items.

The information types supported in the prototype system includes e-mail, to-

dos, notes, bookmarks, music, et cetera. The research prototype allows multiple different organizations of these information types, and disparate information entities may be organized together into collections. For instance a collection may relate to a project, which contain project related e-mails, notes, documents, schedules, and address information for project participants.

Associating descriptive attributes is optional, which means that information may be unstructured, semi-structured, or belong in a structure. This structure may be a collection of information entities, or a list of information entities, or belong in a hierarchy. However, the current prototype interface only has limited interaction support for creation of different information structures, as the focus has been on the concept of unifying interaction in the user interface. Future work includes further development of the research prototype.

Adding new or changing existing associated attributes allows the structures in which information entities belong to change and evolve over time. For instance an information entity may be included in new collections and structures if needed.

Finally the research prototype includes search facilities. It enables searching both the content of information entities and the attributes associated with the information entities. As multiple attributes may be associated with information entities it supports the human ability to recall along different dimensions about the personal information.

1.6 Related Work

This section provides a short summary of related work. The focus is on interactive systems, which provide some level of support for personal information management, or is otherwise relevant to the NEXUS approach. Related work is found both within the research field of human computer interaction, and also in commercial projects or products. The focus is on related research projects, but a couple of commercial products are included in the discussion, as they are highly relevant or widely used. As this section is a short summary only approaches representative of related work is mentioned. Chapter 4 provides a more detailed discussion of related work. Chapter 10 provides a summary of the projects most relevant and contrast them against the approach presented here.

In several projects related to personal information management similar problems have been identified, and similar questions asked. However, different projects have provided very different solutions, each addressing different aspects of the problems.

The most widespread and widely used computer user interface metaphor today is the desktop metaphor. Originally a project at Xerox Parc [JRV+89], but the model has been adopted in commercial products from Apple, IBM, Microsoft, and others. Current implementations have a desktop, which is a spatial layout of information (files or documents). This can be used as a temporary placeholder, for information that is important, or information serving as reminders of tasks to carry out. However, multiple open application windows may hide information placed on the desktop, and they may also clutter the screen. Modern operating systems, which implement the desktop metaphor uses a hierarchical structured file system. This forces a hierarchical structure of user files and documents, and also that the user provides a name for each file. This allows for one organization of information only. Some systems have facilities, which allow the user to create links to files elsewhere in the hierarchical structure, thus simulating multiple organizations of information. Files are handled by different applications depending on the type of information, thus introducing several different modes of operation [Ras00]. Finally, individual desktop applications may of course have application-specific information structures and organization of information

PAD++ is a research system from University of Maryland and University of California, San Diego [BH94], which is a so-called Zooming User Interface (ZUI) [PF93]. It is an alternative approach to computer user interfaces such as the desktop metaphor. The approach is based on spatial organization of information. Information is placed onto a two-dimensional surface similar to a desktop. However, the surface is infinite, and the system allows the user to zoom in to get further detail and zoom out to get overview. Therefore such systems are also known as multiscale user interfaces. PAD++ has been applied for personal information management, where the personal information is simply organized in a spatial way on the infinite surface. The idea is also that naming information entities is not necessary. Obviously this approach may support less structured personal information when placed onto the surface in no particular structure similar to organization of papers on a physical desk. Additionally it may support reminding of tasks to carry out, simply by leaving information in

specific locations on the surface.

Lifestreams is a system from Yale University, which is also an alternative to the desktop metaphor [FFG96b]. Lifestreams is based on a *diary* metaphor, that is, an organization of files and documents into temporally organized streams. Information may include documents and files, but also e-mails and calendar events [Fre97], thus not all information needs to be handled by individual applications. Information is automatically organized on to the temporal stream, which means that the user is not forced to specify a specific name or other identification of information entities.

Placeless Documents is a research project from Xerox Parc, which aims to facilitate document management, by addressing the problem of hierarchical organization of information [DELS99]. The project is not aimed at personal information management specifically, but document management in general. In this approach files and documents are tagged with a set of descriptive attributes (metadata). The attributes are both system derived and user defined. The user can add any number of attributes to files and documents. Additionally the documents are placed onto a surface, and associated attributes are shown visually next to the documents. The system also allows grouping of sets of documents. Placeless Documents may facilitate some aspects of personal information management, but as personal information management was not the focus of the project it has not been studied.

Haystack is a research project from MIT [AKA99] similar to the Placeless Documents project. One of the aims of Haystack is to facilitate personal information management specifically. Similar to the Placeless Documents approach information have associations to descriptive attributes (metadata). This allows multiple organizations of information entities. Information is displayed in views, and information entities may be viewed and accessible from within multiple views. Thus an information entity may be used in multiple views, which may be used for different tasks and in different context of use. The prototype user interface has a display that contains a set of views simultaneously. For instance four different views containing e-mails, calendar, news items, and items related to a project. As all information are located in these it is unclear to what extent the approach facilitates the reminders. The user interface of the Haystack prototype has navigation similar to hypermedia (webbrowser) navigation.

Lotus Agenda is a commercial text-based DOS software product from Lotus Development from the late 1980's, which aim to facilitate personal information management [KKB+90]. Unlike other personal information management software applications Agenda is based on a simple item-attribute-view model similar to the model mentioned above. An information item is a short string of text, with the possibility of associating descriptive text. Attributes can be added to the information items. Thereby the information items are organized into different views depending on the attributes associated. This allows for multiple organizations of the same information. Even though Agenda has a text-based user interface, the system has many aspects of a unified user interface and interaction like NEXUS, as common commands apply the same in the system independent of an information entity representing an appointment or an address. Attributes in Agenda are a little different, as they belong in a hierarchical structure.

There are multiple ways of visually presenting information [Shn96]. As personal information management involves different types of activities and different types of information, appropriate visual presentations of the personal information may depend on context and the purpose of use. The research prototype user interface allows the user to choose among multiple presentations of personal information, and to create unlimited additional views of personal information. Thereby using a view of information, which is useful for the task and context of use.

The purpose of the NEXUS approach and the approaches mentioned above are to support and facilitate the different processes and information types involved in personal information management. The NEXUS approach has elements, which are similar to the approaches mentioned above. However, there are no studies, which compare the approaches above in terms of supporting and facilitating the different aspects of personal information management. This leads to a comparison of the approaches as such, rather than their individual applicability for supporting and facilitating personal information management.

The NEXUS model is similar to the one used in Agenda, and partly in Haystack. The NEXUS approach emphasizes the aspect of providing unified interaction and user interface across different types of information and different organizations of information. This is only partially supported in the Agenda and Haystack approaches. Additionally it supports multiple ways of organizing

information, and does not rely on a single type of organization only, such as a, hierarchical, spatial, or temporal organization approach. Some other approaches allow multiple organizations of information entities too. As mentioned the purpose of the NEXUS approach is to support and facilitate the processes involved in personal information management, however it is unclear how it compares to the other approaches mentioned. For instance it is unclear how well reminding is supported by the NEXUS research prototype.

The different approaches to support personal information management are discussed in further details in Chapter 4 and Chapter 10 summarizes the approaches closest to this approach and they are contrasted against the NEXUS approach.

1.7 Contributions

The contributions of this thesis are considered to include:

- An attribute-oriented model on which personal information management systems may be built, enabling unified access and interaction with information in the system
- A research prototype implementation that serves as a proof-of-concept of the model for personal information management and a test bed for further experiments
- Experiment involving semi-structured interviews of subjects in order to get an understanding of present practices and problems in personal information management
- Pilot usability experiment with the research prototype implementation evaluating the approach in terms of providing support of personal information management
- A survey of prior personal information management studies and an identification and clarification of the issues involved in personal information management

- A comparative analysis of different approaches to the support of personal information management (in interactive systems)
- A set of directions for further research in the area of personal information management
- Identification of issues in personal information management that may inform designers of issues to consider in the design of such system

1.8 Outline

The outline of the thesis is as follows.

Chapter 2 introduces the concept of Personal Information Management. This introduction includes a discussion of studies of how people perform personal information management, and the common findings that has been generalized from research in the area.

Chapter 3 discusses cognitive aspects of personal information management. This includes issues such as recognition and recall, that is, specifically the role of memory in personal information management.

Chapter 4 discusses the support of personal information management in interactive systems. It includes a presentation of different approaches and models to the support personal information management in interactive systems. The advantages and disadvantages of the different approaches and models are discussed in terms of specific personal information management activities and cognitive issues.

Chapter 5 further discusses problems in current approaches from an interface and interaction point of view. A set of specific issues is discussed, and it is discussed how personal information management may be supported and facilitated through a unified user interface and unified interaction.

Chapter 6 discuss models for personal information management systems on three levels – conceptual models, interaction models, and data models. An attribute-oriented model for the information storage is presented and discussed in terms of personal information management support. This includes a discussion of the components of the model and specific issues. Additionally it is discussed

how the model may serve as a foundation for a unified approach to the support of personal information management.

Chapter 7 describes the implementation of a research prototype, which serves as a proof of concept. This includes a presentation of the implementation environment chosen and a description of the implementation of the application of the model for personal information management. The user interface of the research prototype system is presented and discussed.

Chapter 8 describes initial experiments conducted using the research prototype. The experiment included interviews with test subjects investigating personal information management habits and practices. This was followed by a pilot usability experiment where test subjects were asked to carry out a set of simulated personal information management related tasks using the research prototype.

Chapter 9 presents the results of the experiments, and discusses these results in relation to the system model and personal information management issues presented in Chapter 2. This includes a discussion of the support for personal information management using the model.

Chapter 10 provides a summary overview of related work addressing specifically the support of personal information management in interactive systems. Using this discussion of related work it is possible to contrast the NEXUS approach against the others.

Chapter 11 contains conclusions, a summary of contributions, and directions for future work.

Chapter 2

Personal Information Management

This chapter introduces personal information management in general. Personal information management has been defined by Lansdale as: "The methods by which we handle, categorize and retrieve information on a day-to-day basis" [Lan86b] and by Thomas and Jones: "[Personal information management] refers to the methods by which individuals handle, categorize and retrieve information" [JT97].

In the following sections the different terms are discussed further including the different aspects of personal information management. This is based on studies of personal information management behaviour in different domains. Studies of behaviour and personal information management use have been conducted in both traditional office environments and by means of interactive systems supporting personal information management. The discussion in this chapter includes the types of information and activities, which have been identified in these studies of personal information management.

The issues relevant in the support of personal information management are emphasized. That means potential implications for the design of systems, which aim to support and facilitate personal information management.

2.1 Overview

Humans keep and manage personal information for different purposes as part of both their private and professional lives. However, as human memory capabilities – both short term and long-term memory – are limited [Mil56] tools are used, which support our memory and the process of information management.

2.1.1 Personal

Personal information is information that is kept for individual future use and reference based on individual criteria. Thus a collection of personal information may be meaningful to the individual only and in the context in which the personal information is used and understood by the individual.

When information is personal it might also be private, meaning to be seen and handled by an individual only. For instance personal correspondence is typically considered to be private and for the individual to be seen only. Thus the topic of personal information management includes a set of privacy related issues. The privacy issue has a set of implications for the management of the personal information and it also implies a set of implications for the design of computer-based systems to ensure that information is available to the individual only. However, even though the privacy issue is a highly relevant aspect of personal information management, it is not the focus of this project. Whether a piece of personal information is private or not is not a concern in this project.

Rather the focus is on the information being personal as such. Personal information includes both private information and public information. A simple example can clarify the distinction between the two. E-mails are typically considered private information, especially if it involves personal correspondence. A bookmark to a webpage, that is, the reference as such is not private as such but public. Seen, as an isolated piece of information, *the reference itself* is not a private piece of information. However, a bookmark may still be a piece of personal information if it is stored in a collection of bookmarks by an individual. Seen in the context of its use the bookmark might be considered to be private information. For instance the collection of bookmarks may be considered private. In any case the bookmark is personal information in the sense that it represents a

personal interest to the individual.

Therefore personal information is either private information or it represents a personal interest in information that is not necessarily private. Thus information being personal means that personal information can be characterized as either being private information or having individual interest.

2.1.2 Information

Having discussed the *personal* aspect in personal information, the *information* as such is discussed.

With the definition provided in the previous section basically any information can be characterized as being personal information. Strictly following the definition implies that in case an individual has an interest in information and wants to keep that information or a reference to that information it is characterized as being personal information. In order to clarify and delimit the types of information, the focus is on information types that are typically addressed in personal information management systems and in studies of personal information management [BN95] [ABC98] [Mac88] [WS96].

In a sense personal skills such as information on how to shift gear in a car or typing on a keyboard may be considered personal information at some level. However, this low level type of information is internalized and automatic, and takes place without conscious attention or control. Therefore the *information* does not have a representation outside a human being. It is even difficult for humans to express the sequence of actions involved in driving a car once those actions have become learned and habituated – a skill [Ras83].

The focus is on information that is explicit and can be expressed by the human, so that it may be stored in an external system using a suitable symbolic representation. Personal information may include a wide variety of different types of information, which is used as part of different daily activities in everyday life and/or professional life.

Information that may be characterized as personal information includes, but is not limited to, the following types of information:

- Addresses and phone numbers
- Calendar and schedules
- Appointments
- To-do items (reminders of things to do)
- Messages (paper memos, e-mail messages, or instant messages)
- Bookmarks and other types of references to information
- Facts and ideas
- Notes and annotations
- Journals and diaries
- Invoices and book keeping information

The information types mentioned above are typically used in relation to a diversity of activities. Different types of personal information may be used together as part of those activities. For instance scheduling a meeting may involve scheduling information, calendar information, participants contact information, and perhaps messages to the participants.

2.1.3 Management

The *management* of personal information can be seen as an activity carried out by humans in order to keep their personal information for later use. These activities can be divided into two different types of activities.

- Activities in which personal information is relevant and is used. In this case the personal information is not the primary focus. Rather the focus is on the higher-level activity in itself, and the personal information is only used as part of that activity in order to obtain the goal of the activity. For instance information such as, e-mails, reminders, and contacts may be used as part of ongoing work activities, as mentioned above. However, that information only serves the purpose of achieving a higher-level goal.

- A subset of activities is activities where the primary focus is the personal information as such. In that case the management of personal information is an activity in itself. Management of the personal information means the collection and handling of the information, which includes tasks such as, storing, categorizing, organizing, and discarding of the information.

An individual keeps personal information for future reference and use. This involves management of the personal information as a necessary process in order to ensure that the information can be found and re-used. Such management of personal information may be approached in different ways using different strategies. As a simple example of such an approach is information, which is stored in a specific location in order to enable later easy retrieval and use. For instance it may be a note left on the desk as a reminder of a task to carry out. The information is the note with a reminder, which is kept and handled by placing it in a particular physical location, which enables easy retrieval and use of the information.

Personal information management includes the methods used by individuals to collect and handle information. This involves some organization of the information, which enables retrieval of the information at some point later in time. As such this is similar to the traditional problems addressed in information retrieval systems and research, that is, storage and retrieval of information. Therefore

management of personal information involves a set of the same problems found in those types of systems.

However, *personal* information management differentiates from information management in general in terms of several aspects. The previous two sections already mentioned the *personal* aspects of the information, and the different types of information typically addressed.

Some specific aspects which differentiates personal information management include:

- The information is *personal*, and thus may only be meaningful to an individual
- Information is kept for future reference and use, based on *individual criteria*
- When retrieving information, the user has already handled that information (chosen to keep it and store it). Thus information kept by the individual is *known* by the individual
- Humans have *memory* or partial memory of all personal information that is kept
- The *structure of the information is created by the individual* (not by a system), and thus may only make sense to that individual (in the context of its use)
- The organization and structure of the *information reflects an individual understanding (knowledge) and use of the information*, and may only be meaningful to the individual in a given context known by the individual

Some activities are also specific to personal information management, as mentioned in the previous section. In information management activities the personal information is the primary focus. The different types of personal information identified can be related to different personal information management activities.

Examples of personal information management activities include:

- Scheduling, e.g. keeping appointments
- Task management, e.g. keeping reminders of tasks to do
- Communication, e.g. using and managing e-mail messages (or instant messages)
- Contacts management, e.g. keeping addresses and phone numbers of contacts
- Keeping references, e.g. bookmarks to webpages or other information items of interest

As mentioned in the previous section these information management activities may be a part of activities at a higher level, thus serving the function of achieving a goal at a higher level. Therefore such a higher-level activity may involve and combine several of the activities mentioned above.

A key issue in the management of personal information is how the information is organized. The organization is essential to enable retrieval of the information is to be easy and efficient. These issues have been identified in studies of personal information management behavior.

2.2 Studies of Personal Information Management

A number of studies of personal information management have been carried out using different methodology. Typically interviews and questionnaires have been used [Mal83] [Col82] [BN95] [FFG96a] [Fre97], which implies asking a set of subjects about their personal information management behaviors. Results from such studies of course have to take into account, the uncertainty in terms of the gap between what humans say they do and what they actually do. Methods also include observations, surveys [Col82], analysis of individual information structures and use, that is, studies of personal information management of different types of personal information including, files [BN95], e-mails [Mac88] [WS96],

bookmarks [ABC98], calendar [KC82]. Finally analysis of logs from personal information management systems actually used by humans to perform personal information management tasks has been applied also [DCC+03].

Some studies have focused on archived information, that is, retrieval of information from an archive system, but research suggests that personal information management is a more complex issue than archived information. As such, the studies are not comprehensive and do not thoroughly cover all aspect of personal information management [WTN00]. More research is needed to establish a further and more detailed understanding of the complexity and all aspects involved in personal information management.

Nevertheless, a set of common aspects of personal information management behavior has been identified. This behavior of humans performing personal information management has been identified across different domains of information types, applications, and use.

Personal information management was done in the physical office space prior to the introduction of computer-based systems. Studies have been carried out both in the traditional paper-based office domain and in interactive systems that support personal information management. Malone [Mal83] and Cole [Col82] have reported findings of how people organize and use their personal information in office environments.

Malone [Mal83] reported how people handled information and organized their desks in order to manage personal information and carry out daily work tasks from his classical study of 10 office users. It was based on interviews with professional and clerical office workers. Based on the survey Malone presents a set of suggestions and directions for the design of interactive personal information management systems. The study also identifies a set of requirements for interactive personal information management systems. Cole [Col82] also studied human aspects of office filing. Based on structured interviews with 30 subjects, results similar to those of Malone were obtained. Cole also identified some requirements for interactive systems.

Similar studies have been conducted in the interactive systems domain to study the practices in electronic filing using different computer operating systems. The purpose was to find out how people handle and organize personal information using interactive systems. Barreau and Nardi [BN95] did several stud-

ies of how people do personal information management using computer desktop systems. Subjects used different systems; DOS, Windows, OS/2, and Macintosh, but a set of common findings were reported. They found similar work practices among users of different computer-based personal information management systems. It is noticeable that these studies report findings similar to the ones found in the physical domain studies by Malone [Mal83] and Cole [Col82].

Research of personal information management has also included studies in different domains of use. First there is the set of studies of paper-based personal information management [Mal83] [Col82] [KC82], and a set of studies of use of different types of personal information facilitated by interactive personal information management systems. The latter includes studies of file management [BN95], e-mail management [Mac88] [WS96], and bookmark management [ABC98] [JBD01]. Some common findings have been reported from these studies of different personal information management use in different domains.

In the following sections these common findings from multiple studies of personal information management are discussed.

2.3 Individuality

An important finding is the high degree of individuality of personal information management needs, usage, style, strategies, and habits. These are highly individual and therefore widely varying.

2.3.1 Needs and Usage

Humans are fundamentally different in their needs for personal information management. One aspect is the amount of information managed by individuals, which represents a wide spectrum. Some keep very little information, and will find it to be efficient with no systems or tool as such, or simply a traditional paper-based calendar or notebook. Others use and keep extensive quantities of personal information. For instance because they send and receive a considerable number of e-mails and/or keep information from many distributed information sources.

The study by Malone indicates that type of job may have some influence on the management behavior [Mal83]. Other aspects of individuality are which information types are kept and used, how frequently, and in which situations individuals manage and use personal information. This depends on aspects such as job type and individual approach to information management.

Jones and Thomas report from a simple survey of 23 subjects from organizations in the South West England [JT97]. It was identified that the most widely used artifact was the "to-do" list, which were used by 60% of the subjects. About 50% of the subjects were using calendars and address books, and between 30 and 40% were using pocket diaries, desk diaries, and personal organizers. Only 9% of the subjects were using electronic organizers and personal information management software. Thus only a modest adoption of electronic personal information management technologies was found. The subjects in the study who did employ electronic tools typically used a combination of paper-based and electronic tools for personal information management. However, these results are limited in two ways. First, the sampling does not allow the results to be scaled to a wider population. Subjects were drawn via accidental/opportunity sampling. Second, the use of the World Wide Web, e-mail, and Personal Digital Assistants has changed considerably and become widespread. This indicates increasing demand for *electronic* based personal information management, which other studies identifies [WS96] [ABC98] [JBD01]. Nevertheless, the study does indicate the wide variety in personal information management needs and usage.

2.3.2 Individual Style, Strategy, and Habits

Another aspect of individuality is the style and strategy chosen by individuals to manage the personal information that is kept. As the information is personal and kept for personal use only it is to be retrieved by the individual only. The individual chooses a strategy that works from an individual perspective and is *perceived* by the individual as being efficient. This has implications for the styles of management and organization of the information chosen by individuals. The strategies used for keeping and managing the information is widely different.

Consequently humans develop and habituate individual strategies for the management of their personal information. There are different behaviours and

styles, which means that these strategies are highly individual. Examples of personal information management strategies include building piles of documents in an office environment, or creating and maintaining an extensive hierarchical filing structure in an interactive information system. Such individual management strategies are sufficient as the information is personal, and therefore is only retrieved by the individual. This means that the individual must make use of his own personal strategy or system in order to retrieve and re-use his personal information.

In the study by Malone [Mal83] two general office organization styles were identified. Those styles are characterized as the "neat" office and the "messy" office. The neat office is characterised by having only a few structured piles (e. g. in- and out-basket), and the documents are organized by some systematic scheme. Also only small amounts of documents were observed on the desk. The messy office is characterised by having more unstructured piles, and the documents are not filed. Instead they are placed into "loosely stacked" piles. The desks observed had many unorganised papers and documents on the desk. A similar distinction was identified and reported in a study by Lansdale [Lan88]. Also Mackay observed that some subjects were very organized while others were very disorganized in a study of e-mail use [Mac88].

This division of management styles is coarse, and only provides a rough characterization of individual behavior and style. However, it does indicate that individual differences in filing style does play a role in the management of personal information and illustrates the variety of individual approaches to personal information management. Individuals with the "neat" office tended to use standardized and rigid procedures, whereas others with the "messy" office use informal or ad hoc rules of filing. As such there is an important element of *individual style* to personal information management to consider.

A similar way of charactering these different types of styles is the strategy used to keep and handle information. The individual organization styles can also be characterized as "filing" and "piling". Filers "file" their information, whereas pilers "pile" their information, corresponding to the "neat" and "messy" offices. Cole [Col82] also reported that "filing systems were not rigidly maintained; rather organisation was idiosyncratic, dependent upon the demands of the users and their motivations". This distinction is also mentioned by Jones

and Thomas [JT97].

Any individual strategy may be characterized anywhere in between these two characteristic styles, for instance a mix of these approaches. These approaches characterize the way humans deal with personal information and develop their own information managing style and habits.

The individual style and strategy for information storage was also identified in a study of how subjects kept found information on the web [JBD01]. This study also reported a great diversity in the methods used by individuals to organize information for later use. The subjects were both flexible and creative in terms of developing methods and strategies for information storage and in their practices of information re-use.

2.3.3 Challenges for Information Management Systems

As personal information management is highly individual, this poses a set of challenges in the design of personal information management systems. Especially if the tools or systems aim to support every need, as the needs, behavior, styles, and habits varies considerably. The issue is the possible *scalability* of a particular approach in terms of information quantities, types, frequency of use, and situations of use. Some humans choose to do close to no personal information management, and thus hardly need a tool or system at all. Other segments of humans have a little to medium need for tools and systems as they only handle limited amounts and types of personal information. Groups that have a medium or large use of personal information management may need tools and systems in order to handle the information in an efficient way. That is to allow easy retrieval and re-use of the information.

The focus of this project is on the latter groups, which have a high demand for personal information. This means, that they may have a need for interactive systems in order to facilitate and support the processes involved in collecting and handling large quantities of personal information, which is used as part of their everyday activities.

2.4 Reminding

An interesting finding is that reminding is a very important feature in personal information management. Subjects in the study by Malone [Mal83] left things in specific locations on the desk and tables reminding them of things to do. Malone concludes that it is important that computer-based information systems serve this reminding function as well, and that the information should appear automatically without being requested. The physical location of documents itself served as the reminder of things to do why the subject did not have to take explicit action in order to be reminded. Therefore the reminding function in the office arrangement and information organization in the office is an implicit function, meaning that it does not have to be explicitly invoked [Nie93a].

Barreau and Nardi also identified the importance of the reminding function in their studies [BN95] [FFG96a]. Even if interactive systems do not directly support the reminding feature, users of such systems simply create ad hoc methods to remind themselves of task to do. For instance leaving icons in a particular places on the desktop served as the reminding function. Again a particular location of a file itself served as the reminder of things to do. The reminding feature thus becomes an implicit function similar to the approach used in the physical environment as found in the study by Malone. Barreau and Nardi point out how the reminding feature is closely related to the fact that the subjects in their study preferred location-based finding. The personal information management system must therefore allow the user to leave reminders in locations where it will be noticed, such as, on the desktop or in a specific file folder.

The importance of reminding was also indicated in a study of the use of bookmarks in webbrowsers [JBD01]. It was observed how reminding was obtained by using different ad hoc strategies. For instance subjects in the study would send and e-mail containing a bookmark to themselves in order to keep a reminder to retrieve a given webpage at some point later in time. The standard bookmark feature in the webbrowers did not support the function of reminding. Thus the subjects had to use ad hoc strategies to cope with the lack of support for reminding as such in the system.

Generally leaving e-mails in the inbox of e-mail systems also serves as reminders. Even though e-mail systems are designed to support asynchronous

communication, studies have shown that humans create coping strategies due to the lack of support in the system. The e-mail inbox is a place (a location), which is frequently visited and checked by the users and therefore may serve the function as a location to keep reminders [Mac88] [WS96]. This observation has lead to recent studies and research of systems with user interfaces, which have a tighter integration between e-mail and task management [BDHS03].

Finally it should be noted that the distinction between finding information and being reminding about information might be unclear. When finding a piece of information that was intentionally looked for, it can be characterized as *finding*. However, if a person becomes aware of a piece of information unintentionally while performing another task it can be characterized as *reminding*. For instance information left on the desk or in piles reminds individuals without having to rely on explicit searching [Mal83].

2.5 Information Discovery, Storage, and Retrieval

As discussed in Section 2.3 personal information management is highly individual, and therefore individual strategies for managing information is chosen. Managing the personal information includes the processes of collecting, handling, and keeping the information. This involves some level of organization of the information depending on individual style and information use. However, developing efficient strategies for organizing the information, which enables efficient retrieval and re-use of the personal information are not trivial.

Individual strategies and styles are used as identified in the studies by Malone [Mal83], Cole [Col82], Mackay [Mac88], Barreau and Nardi [BN95] and others. However also a set of common findings were identified in their studies. Independently of the individual differences there is a set of common findings for unit of information structures and the creation of information structures for personal information.

2.5.1 Information Discovery

Information management involves the processes of storing and retrieving the information. However, the very first step is of course to obtain the information that potentially will be kept by the individual for future use. Only at that point the information has become *personal information* in terms of the definition provided in the beginning of this chapter. This initial step is referred to as *information discovery*.

Information may be obtained or discovered by an individual in many different ways or arrive through different channels. A few examples of information discovery includes:

- An individual might get an idea that is converted into a symbolic representation, that is, information that represents the idea
- Information found while intentionally searching information
- Information might be discovered unintentionally while performing another task
- Facts that are obtained
- Information received by the individual (through whatever medium or channel)

This means information discovery may origin from multiple sources.

Once a piece of information has been discovered it is judged by individual criteria whether it should be kept. The intention is that there is possible future re-use of the information. Thus the information or a reference to the information is kept. The overall purpose is to enable the individual to retrieve and re-use the information easily and efficiently.

2.5.2 Storage and Retrieval

Storing the information or a reference to the information saves the individual from having to repeat going through the sequence of steps that was originally

performed in order to discover the information. In some cases this sequence of steps might be long and tedious as it was difficult to discover the information initially. Whereas in other cases only few steps was necessary to discover the information. Even though information might be easily *discovered*, it may still be a difficult issue performing the storage of that information for later use. For instance receiving information is easy "information discovery", but storing the information for later use is still a non-trivial issue.

The challenges in information storage and retrieval systems is the difficulty of storing information in the systems, and the difficulty of retrieving the information from the system. The retrieval of information should obviously be less difficult than repeating the sequence of steps used to discover the information originally [JBD01]. Otherwise the storage-retrieval system serves no purpose at all.

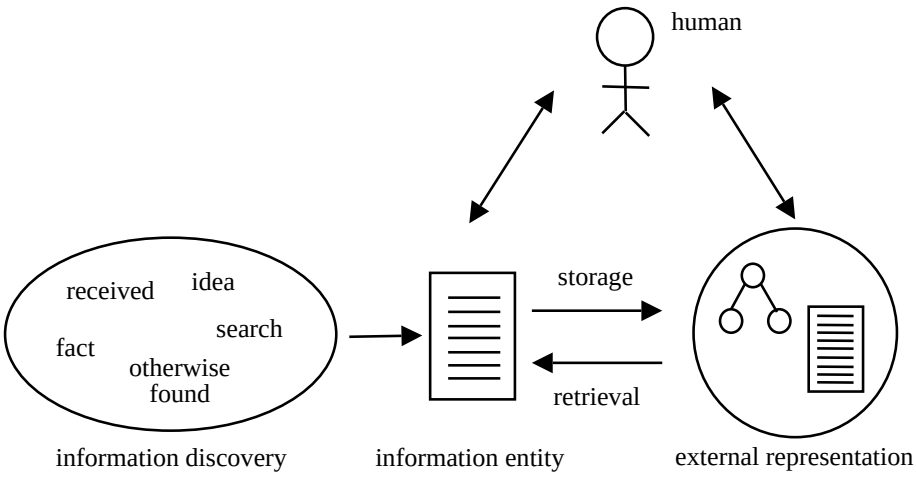

Figure 2.1: Illustration of the processes of discovery, storage, and retrieval of information.

An information item is discovered, and then stored and retrieved from a system, which contain the personal information.

Figure 2.1 illustrates the processes of discovering, storing, and retrieving information. The initial step is the information discovery, which may originate from different sources. The discovery of the information as such is less relevant for personal information management. The information is not considered personal until after the information discovery process has been performed and the individual has decided to keep the information. Therefore personal information management in terms of the definition provided in the beginning of this chapter only includes the processes illustrated after the information discovery process has taken place.

At that point the individual decides what to do with the information that was discovered. If it is decided that the information should be kept for later use it is stored in a system based on an individual strategy. Later it can be retrieved from the system by means of the individual strategy that was used to store the information.

2.6 Information Structures

As mentioned in Section 2.3 the individual strategies chosen for managing and keeping personal information varies. This also means that the level of formality of information structures varies considerably.

For small volumes of information an ad hoc filing strategy might be sufficient to enable possible later retrieval and re-use. Such ad hoc strategy might be to put information in the nearby office environment or a simple pile on the desk. As the number of information items in the pile is relatively small, it is fairly easy to go through the pile to retrieve the information needed. However, as the quantity of information increases, this approach becomes less efficient. Larger volumes of information mean that it must be organized using a more systematic approach in order to allow retrieval and re-use [Col82].

In several different studies of personal information management similar characteristics of information management strategies and information structures have been identified. These studies include [Mal83], [Col82], [BN95], [Lan86b], and [Lan88]. This has been observed across domains, both from studies of information management in office environment, and by means of interactive information

management systems. The different types of information structures that has been identified in multiple studies of personal information management include:

- Unstructured information
- Semi-structured or informally structured information
- Formal information structures

However, these observations of differences in the formality of information structures can be attributed to different aspects of personal information management.

One aspect is the *individual differences* in the way individuals manage their information. The strategies chosen by individuals vary. Another aspect is the *differences in information types and use*. Not all personal information managed by an individual is handled the same way using one type of information structure only.

2.6.1 Individual Differences

First, individual differences in information management strategies contribute to differences in the formality of information structures. As discussed in Section 2.3.2 a rough characterization of information management behavior is those who *file* information and those who *pile* up information. This corresponds to the neat and messy offices characteristics identified [Mal83].

This implies that the personal style and strategy has an effect on the way information is structured, that is, to which extent the personal information is structured. Some individuals prefer to keep information in semi-structured piles and retrieve information by searching through those piles. Others prefer formally structured information organization and retrieve information by means of the formal information structure created.

2.6.2 Differences in Information Types and Use

Second, differences in information types and use also contribute to different information structures. This means that an individual might also use the whole

spectrum of information structures for different subsets of the personal information.

A subset of the information may be unstructured. Some information is kept on an ad-hoc basis, using ad hoc strategies for organizing the information. For instance just leaving it temporarily in the immediate environment. Such information is typically not structured, as the perceived effort to do so would exceed the value of keeping the information in a structured way.

Information may also have an informal structure or be semi-structured. An example of a strategy for keeping information this way is to keep information in piles as mentioned previously. In the study by Malone [Mal83] it was identified that piles have no systematic order, but their spatial location is often especially important. Even though piles may seem to have no information structure at all, studies have identified, that they might have an implicit order. For instance the newest information item may always be placed on the top of the pile or at least near the top of the pile. The further down the pile the older the information items are. Therefore such piles have an implicit temporal order. The piles themselves are put in different spatial locations in the environment. Each may serve different purposes and perhaps contain different categories of information. This leads to a characterization of piles as having an informal information structure.

Finally, there are formal information structures, which are obtained by using a system to manage and file information. In that case the information is archived using a particular filing scheme, so that information can be retrieved by means of the scheme in a systematic way. This is relevant for personal information, which is kept for a longer time, which will be discussed in the following sections.

The differences in information structures imply that personal information can also be characterized by the extent in which it is structured.

2.7 Classes of Information

In the study by Malone [Mal83] it was also observed that the frequency of use of personal information is varied. He distinguishes between information that is used frequently that must be easily accessible, and information that is archived for later retrieval. The latter is associated with the problems of properly

organizing and categorizing the information so that the retrieval is efficient. Malone distinguishes between three different classes of information identified in file organization and use of paper documents in offices:

- Frequently used – active information
- Reminding – things to do in the near future
- Archived information – archived for later retrieval

A similar classification of information characteristics was identified in other studies. Cole [Col82] describes information characteristics using three categories. Also in the studies of users performing file organization using different desktop systems carried out by Barreau and Nardi [BN95] information characteristics is described using similar three categories.

As mentioned in Section 2.4 these studies have also identified the importance of reminding. The information can be characterized using the following three categories.

- Ephemeral (action information) – short life information
- Working (personal work files) – frequently used information
- Archived (long term storage) – infrequently used information

In terms of personal information management these categories represent different levels of interaction, that is, handling and use of the information. Also the different types of information are characterized by the frequency of use and lifetime of the information.

The studies by Malone, Barreau and Nardi are mainly based on interviews and demonstration of personal information management practices. Thus interviewees provide a description of their personal information management practices in retrospect. As such the studies are not actual observations of use and behavior. The subjects describe their perceived behavior. There might be a difference between what they do and what the actually do, but it is unclear whether the difference is significant for the results. However, as described above various studies have found similar classifications to be suitable for describing information handling and use.

2.7.1 Ephemeral Information

Ephemeral information is also referred to as "action information". Information that is used at this level of interaction is characterized by a short lifetime. This means the information is used or going to be used in the near future. When the information has been used it may not be needed for future use and might be erased. Such information may include documents, lists, e-mail messages, notes, or to do items. For instance an e-mail may be received and only kept for a limited time and used in the near future. As it may not be needed for further reference it is not kept, but deleted immediately after its use.

In the study by Cole [Col82] this type of information was identified as being immediately accessible in the nearby environment. For instance it might be located in in-trays, on the desk, the floor, or in piles.

The accessibility of ephemeral information is gained by keeping the information nearby in an ad hoc manner. No or very little systematic organization of the information is performed. This means that ephemeral information is only efficiently used if the quantity of information is limited. A small quantity of ephemeral information allows it to be located based on remembering the information or the location, or simply by its presence in the nearby environment. Larger quantities of such unorganized information mean that information items may have to be manually searched for, which leads to less efficient use of such information.

Barreau and Nardi [BN95] observe that the "central problem of organizing ephemeral information concerns where and how to file information that is needed for only a short time". In their study of file organization using desktop systems it was found that the subjects did not have good ways of dealing with ephemeral information. For instance it was kept in top-level directories, or simply left on the desktop. "With limitations on the amount of information that can be viewed on the screen at one time, managing large quantities of ephemeral information can be problematic." [BN95]. In the study quantities of ephemeral information is reported as a particular concern. Files are created for temporary use, but it was reported how they stayed in the system longer than intended. This means that the information ends up having a longer lifetime and therefore becomes something in between ephemeral and working information. "The problems of filing

this information were intensified because the information outlasted the normal shelf life of ephemeral information" [BN95]. When quantities of ephemeral information increases it leads to problems as the information is typically kept in an unstructured way. Especially the support of ephemeral information in interactive systems is an area for further research.

2.7.2 Working Information

Ephemeral information is also referred to as "action information". Information that is used at this level of interaction is characterized by a short lifetime. This means the information is used or going to be used in the near future. When the information has been used it may not be needed for future use and might be erased. Such information may include documents, lists, e-mail messages, notes, or to do items. For instance an e-mail may be received and only kept for a limited time and used in the near future. As it may not be needed for further reference it is not kept, but deleted immediately after its use.

In the study by Cole [Col82] this type of information was identified as being immediately accessible in the nearby environment. For instance it might be located in in-trays, on the desk, the floor, or in piles.

The accessibility of ephemeral information is gained by keeping the information nearby in an ad hoc manner. No or very little systematic organization of the information is performed. This means that ephemeral information is only efficiently used if the quantity of information is limited. A small quantity of ephemeral information allows it to be located based on remembering the information or the location, or simply by its presence in the nearby environment. Larger quantities of such unorganized information mean that information items may have to be manually searched for, which leads to less efficient use of such information.

Barreau and Nardi [BN95] observe that the "central problem of organizing ephemeral information concerns where and how to file information that is needed for only a short time". In their study of file organization using desktop systems it was found that the subjects did not have good ways of dealing with ephemeral information. For instance it was kept in top-level directories, or simply left on the desktop. "With limitations on the amount of information that can be viewed

on the screen at one time, managing large quantities of ephemeral information can be problematic." [BN95]. In the study quantities of ephemeral information is reported as a particular concern. Files are created for temporary use, but it was reported how they stayed in the system longer than intended. This means that the information ends up having a longer lifetime and therefore becomes something in between ephemeral and working information. "The problems of filing this information were intensified because the information outlasted the normal shelf life of ephemeral information" [BN95]. When quantities of ephemeral information increases it leads to problems as the information is typically kept in an unstructured way. Especially the support of ephemeral information in interactive systems is an area for further research.

2.7.3 Archived Information

Archived information is information that is not directly used as part of ongoing activities or work tasks. It is stored and organized for later retrieval and therefore it is less frequently to infrequently used information. Cole [Col82] reports that in the office environment the information is typically stored away from the office as the information is of no direct relevance to ongoing work. In the study by Barreau and Nardi [BN95] the life span of archived information was typically more than a month. Again this is highly task-dependant, and may for instance depend on the lifespan of a project. Archived information identified in the study included completed work, reports, and projects.

The subjects reported some problems with archived information. One of the problems is which files to actually archive in the system and which not to archive. For instance after finishing a project. It needs to be assessed which information may be needed later. One option is to store *all* information related to a project and thereby not having to consider what information to keep and what information to discard. However, there are several problems with this approach. Storing all the information may be a very time consuming and difficult task in itself. As not all information may be needed in the future it is considered to be not worth the effort to store all information. Another issue is that it is perceived as more difficult to retrieve the relevant information among the entire volume of information rather than among a chosen subset of the information.

2.8 Information Organization

The purpose of information organization is to allow easy and efficient retrieval and re-use of the information. The processes involve storing information in a suitable information structure and later retrieving information from that information structure. The previous section discussed how different types of information might be organized with different levels of formality. Also it has been discussed how individual style and strategy may have an impact on the way information is organized.

This section further discusses information organization. The discussion includes issues, which by humans are perceived as barriers to efficient information organization. Both informal and formal organization of information may cause difficulties for humans.

2.8.1 Units of Organization

First an organization of information can be characterized as the information being divided into different units. These units of information can be organized in an information structure, which is more or less formally structured as mentioned in Section 2.7. Independently of information structures these are the *unit-of-work* identified and used by individuals for the organized of information.

	Elements titled	Elements ordered	Groups titled	Groups ordered
Files	Yes	Yes	? *2) maybe*	?
Piles	? *1) no*	No	No	?

Table 2.1: Identification of units of work in desk organization as found by Malone [Mal83]
1) Piles are typically not explicitly titled, but different piles may have overall topics, 2) Groups of files may have a title depending on system

The units of organization may represent different granularity of the information. In the study by Malone [Mal83] the units of organization in the office environment was identified as *files* and *piles*. Table 2.1 shows these units of organization with their different characteristics. Files are characterized by having information elements that have an explicit title, and the information elements are organized in a systematic order. Groups contain a set of information elements, which might or might not have a title and order. Piles are characterized by information elements, which typically do not have an explicit order. However, in some cases piles might have an implicit reverse temporal order. Information elements might or might not have titles. Piles are typically not divided into groups, as a pile itself simply represents one group with not explicit order. Individuals might have several piles, but typically not explicitly titled.

Files and directories where the organization units used by subjects in the study of organization of electronic files by Barreau and Nardi [BN95]. However, these were units made available to the users by the file systems used for organizing their personal information. Similarly both files and directories are characterized by explicit titles and a systematic order. Whereas piles as such were not directly supported by the systems used in the study. Instead Barreau and Nardi observed how subjects used different coping strategies to keep information in "piles" in the system used.

2.8.2 Classification of Information

The mechanism for a more formal organization of information includes creating a classification scheme, classifying the information using the scheme and finally the process of retrieving the information based on the classification scheme. The information is personal and also the information structures and classification of the information is thus personal in the sense that the individual creates it. This means that the approach and strategies are chosen by the individual in order to classify the personal information used as part of everyday activities.

Kwasnik [Kwa89] studied classification of information in the office environment. It was found in the study, that the classification chosen by subjects were person- and situation-centered, rather than being object oriented. That means that information objects were primarily described in terms of their *use* and the

context and relations in which the information is used, rather than an isolated description of the information itself.

Several different dimensions along which information was classified were identified in the study. Examples of dimensions of classifications include: location (spatial property), form (semantic property), time (temporal property), and use (semantic property). Kwasnik reports that *use* was the most frequently used classification for information in the study. Other situation-oriented attributes include source, circumstance, and access. Document specific attributes include author, topic, and form. Another example of information classification dimensions is purpose, urgency, and importance. For reminders dimensions such as urgency and importance are relevant, see Section 2.4.

The findings by Kwasnik illustrate different dimensions of classification that may be used by individuals. This involves categorization of the information by means of the attributes used to describe it. This means that the individual has to create a structure by which the information is organized. However, using a hierarchy to keep information requires human to do something, which have been found uneasy for humans to do well. It has been reported in several studies of personal information management that classification of information in terms of an elaborate hierarchical information structure is problematic.

In the study by Cole [Col82] it was observed that developing elaborate filing schemes was a problem for subjects. It was observed that often humans simply abandon complex filing schemes at some point. Cole reported that there "is a general lack of motivation towards upkeep of elaborate filing systems – in fact, the less time spent filing the better". In some cases subjects reported to have used elaborate filing systems initially. These include color-coding of information and extensive hierarchical classification with many subcategories. However, at some point the subjects reported to have simply given up on such an elaborate filing system for different reasons. "This results in filing systems with all manner of idiosyncracies" [Col82].

In the studies by Malone [Mal83] classifying information was also identified as difficult and causing the subjects difficulties during everyday organization of the office desk. Some of the issues include the "cognitive difficulty of creating appropriate categories and deciding how to classify information in a way that will be easily retrievable" [Mal83]. Also the "mechanical difficulty of creating labeled

file folders, binders and so forth, especially if multiple levels of classification are desired" [Mal83] is reported as everyday classification problems. For some information a hierarchical information structure may simply be seen as being too rigid and difficult for filing of the information. In some cases a hierarchy may not be the natural way for individuals to organize parts of their personal information.

As a consequence the subjects in the study did not necessarily categorize information but rather they used untitled piles of documents arranged by physical location. It was found in the study that accessing information on the basis of the spatial location instead of a logical classification scheme was important. The organization into untitled piles served implicitly as a temporal organization as well. The further down the pile the further back in time the information was located. For instance the temporal aspect were used when subjects were asked to find specific information among the piles in the office. The subjects try to recollect *when* they used the information using the implicit temporal order.

Barreau and Nardi [BN95] also observed that subjects refrained from keeping elaborate filing systems. It seemed that subjects wanted to spend as little time filing as possible. The creation of an efficient archiving scheme was reported as being very difficult, and subjects reported that it was not worth the effort, since it required too much time and effort compared to the value of the actual information. "Every user in the study indicated that their attempts to establish elaborate filing schemes for archived information failed because they proved to require more time and effort than the information was worth" [BN95].

The difficulties of classifying information in itself also lead subjects to defer classification and archiving of the information. Instead information is left on the desk and in various piles in the office, as mentioned in Section 2.3.2.

2.8.3 Storage-Retrieval Dilemma

As mentioned above organizing information is fundamentally about information storage and retrieval. This means storing the information in an information structure so that it can be easily and efficiently retrieved and re-used later. Although this seems trivial studies of personal information management have shown that storing and retrieving information is problematic as mentioned in

the previous section.

As mentioned above, in some studies subjects report that they are reluctant to keep extensive filing schemes. That is elaborate hierarhical structures with many levels and subcategories of categorization. People who have used such schemes report that they skip them after a while. It is perceived as not being worth the effort to keep such extensive schemes [Col82].

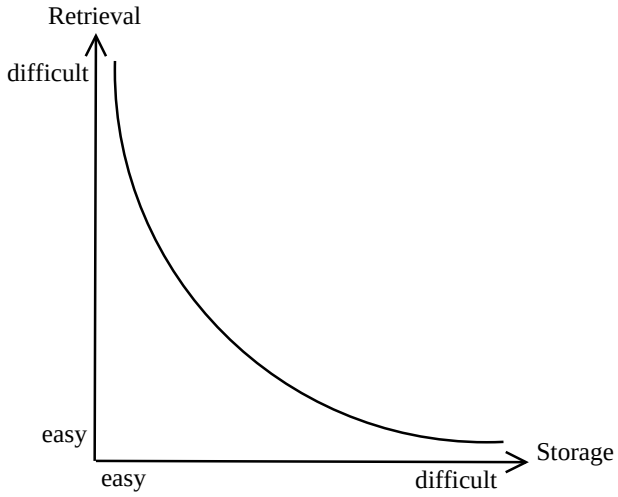

Figure 2.2: The retrieval difficulty as a function of the storage difficulty The shape of the curve is arbitrarily chosen and serves only the purpose of illustrating the relationship between storing and retrieving information as *perceived* by subjects in studies of personal information management

Figure 2.2 illustrates the *perceived* relation between storing and retrieving information in personal information management as reported from studies of personal information management. For instance humans take little storage effort if they leave information in unstructured or semi-structured piles. But some effort might be needed when retrieving information from piles, due to their semi-structured nature. Retrieval in terms of manually searching through piles might be quite difficult and time-consuming. On the other hand systematically organizing information by means of an elaborate information hierarchy might require

some effort and can be perceived as difficult and time-consuming. But then it might enable easy and efficient retrieval of information from the information structure.

The dilemma is that the more difficult it is to store the information the easier it will be to retrieve it again. On the other hand if little effort is spent storing information it may be difficult to retrieve the information later. Studies have indicated that humans want to spend as little time as possible filing information [Col82]. Thus the perceived optimum seems to be to spend as little effort as possible filing information, but at the same time be able to retrieve the information efficiently.

The piling strategy seems to be a reaction to this perceived pattern of information storing and retrieval. Humans develop individual strategies, which seems to enable them to reach a perceived optimum between the time and effort spent storing and retrieving information. It might turn out not to be an actual optimum, but that is difficult to predict, since it may be uncertain whether a piece of information is going to be used later at all. In that case it does not seem worth the effort to spend time storing the information item for easy retrieval at some point later in time. The time seems to be wasted if the information item is never used again. On the other hand *if* the information should be used later, subjects have to cope with spending more time and effort retrieving the information, since little time and effort was spend storing it [Lan88].

2.9 Information Organization Difficulties

As seen in the previous subsections it been observed in several studies of personal information management, that information organization and classification is difficult. Section 2.8 discussed the problems of classification of information as such. However, there are a number of other reasons why humans perceive information organization as a difficult and sometimes time-consuming task. Some of these reasons include:

- Different classes of information is handled differently and needs different information organization
- The understanding of the domain to which the information belongs has an impact on the ease of information organization
- Information use and situations of use changes over time, which might cause information structures to evolve over time
- Multiple information structures may be relevant for the information

The following subsections discuss the different aspects of information organization difficulties.

2.9.1 Difficulties with Different Classes of Information

Barreau and Nardi [BN95] have some interesting reflections about research done in the field of information management. Often this type of research has assumed that people have a significant need for archived information, and various techniques for supporting archived information has been researched. Without a doubt some jobs require people to keep large quantities of archived information, but some jobs do not. In their studies that was true for the managers, administrative assistants, and graphics artists that were among the test subjects. However, the studies of personal information management show that people keep only little archived information, while the quantity of ephemeral and working information is where the primary support is needed. Tools that support ephemeral information have received less attention in research and design. Especially the

quantity of ephemeral information was reported to be a problem in the study done by Barreau. New documents and files are created on the fly for temporary use, but ends up cluttering the user workspace. As such this type of information tends to have a longer life span than ephemeral information, but without being considered working information.

2.9.2 Domain Understanding

Another issue that has implications for the success of organization of information is the level of knowledge and understanding of the domain to which the information belongs. If there is a fairly good domain understanding filing of information can be of little difficulty for humans. For instance an expert within a given domain has a good understanding of the various topics within that domain. This includes an understanding of the structures and relations among information on the topics. Therefore the expert might already have a mental structure of the domain, which can be mapped onto an information structure created in an information system. This makes the process of performing organization of information within a given domain easier for a domain expert compared to a novice.

A novice of a given domain has to develop some level of understanding of the domain initially. Organizing information before the topic is well known is difficult. Until some level of understanding has been developed it might be fairly difficult to perform any organization of information within the domain. The structures and relations among the information are not well understood by the novice. Primarily it is difficult to organize information that is not well understood, but difficulties can also be caused by the potential use of the information being unclear initially.

At least the initial organization of information might lead to information structures that have to be changed later as the understanding of the information develops. Premature organization or categorization of information may cause troubles. When the topics are not well known it is difficult to create a suitable information structure and organize the information within that. This means that initially information may be organized in an unsuitable way, and premature organization of the information may potentially lead to extra effort in terms of

substantial reorganization of the information.

Incrementally the knowledge and understanding of the domain will develop and therefore the capabilities to perform an efficient organization of the information. As the topic becomes better known over time the information structure is likely to evolve as the understanding and knowledge of the topic expands. This means that there might be a need for less formally structured way of organizing the information initially [Mal83]. Otherwise it may lead to substantial reorganizations of the information later, and thus a lot of extra organization effort.

2.9.3 Evolvable Information Structures

Personal information is used over time and as part of different daily activities. Therefore context in which the information is used might change over the lifetime of the information. At some point the information may no longer have any relevance to the activities carried out. This means that the use of the information may change over time. Therefore the structure of the information might also evolve over time. This poses a set of challenges for the creation of information structures and classification of the information.

The reasons for information structures to change over time include:

- As mentioned above the domain knowledge have an influence of the easy of which information may be filed. It might be because the domain knowledge and understanding develops over time, as a topic is better understood. Over time humans becomes more knowledgeable of the information in question. The understanding of the domain expands and the understanding of the relations among the information becomes better. This may cause the human to develop an understanding of the information that has implications for the way the information is organized and structured.
- It might be because the information *use* changes over time. The information may be used in different contexts and situations of use than was originally intended. This means that the structures and filing of the information may evolve over time as the information gains relevance in those different situations of use.

- The relevance of the information may also change over time, meaning that some of the filed information is more frequently used than other information. This may have an impact on information structure as well. Information that is frequently used may change organization so that it is easier to access, whereas information that is never used might be discarded.

The fact that information structures are not necessarily static over time poses additional challenges for systems to support and facilitate personal information management. There are multiple indications the relevance for evolvable information structures for the reasons mentioned above.

2.9.4 Multiple Structures

One possible implication is that multiple ways of organizing the same information may be relevant [Boa01], as mentioned above. This might be relevant as the context in which the information is used changes over time. The information may be used as part of different tasks, why it might naturally belong in multiple categories.

Malone [Mal83] also mentions multiple classifications as important for handling information in his study. "Simply allowing the same document to be easily put in several categories is one way computers can simplify classification". Lansdale [Lan88] also provides arguments for providing multiple keywords in information organization in order to allow multiple categorizations of information. This also addresses the fundamental problem of categorization of information. By allowing multiple categorization humans avoid having to decide under which of two or more categories to file information.

Another possibility is to defer classification to later. When initially using a piece of information it can be difficult to classify the information. Therefore systems could support some way of deferring the classification of the information until classification of the information is possible. In the office environment *piles* has been identified as a way to refrain from initial classification of information [Mal83]. However, the dilemma is that the piling strategy may become counter-effective especially for large quantities of piled information [Lan88]. As such the piling strategy can be seen as not necessarily representing a human need,

but rather as a way to compensate for the difficulties of organizing information in terms of formally classifying it.

2.10 Summary

This chapter has presented the basics of personal information management. This included a description of the various aspects of personal information management, which contribute to its complexity.

From a number of studies of personal information management in different domains and using both paper-based and interactive systems a set of issues relevant for personal information management has been identified. Similar principles for human personal information management have been found whether the information is kept in a paper-based system or in an interactive system [Mal83] [BN95].

There is a *diversity of information types and activities* involved in the processes of personal information management. Individuals manage personal information using their own *personal style* using *individual strategies* for storing and retrieving information. This includes unstructured, semi-structured, and formally structured organization of information. The information and information structures are *not static*, but the information structures and the use of information may evolve over time. Some people refrain from elaborate filing strategies, as the effort to keep an elaborate filing strategy does not seem to be worth the effort. *Reminding* is a very important aspect of personal information management. Personal information management has different levels of interaction, including ephemeral, working, and archived information. These different types of information are typically organized in different ways and have different frequencies of use.

The personal information management issues can be summarized as follows:

- Diversity of information and information types
- Diversity of activities
- Information is related across activities
- Highly individual style and personal information management behavior. This includes and different levels of use, and different strategies in handling the information. For instance filing and piling strategies
- Information structures and organization varies
  - Structuring of information varies from unstructured information, loosely or semi-structured information, and formally structured information
  - Classifying information and creating a filing scheme is a difficult task. It has been observed in studies that often people give up on elaborate filing schemes
  - Storage and retrieval of information is perceived as a dilemma. People want to spend as little time storing and retrieving information, and want both to be efficient
  - Information structures and information use are not static, but evolves over time
- Reminding is an important aspect of personal information management, that is, the ability of being reminded of tasks to do in the future, without intentionally searching for that information

- Different levels of interaction with personal information has been identified [Col82] [Mal83] [BN95]. This means information is typically handled and structured in different ways:
  - Ephemeral information – is "action information". This means that it is characterized by a short lifetime and it is used in the near future.
  - Working information – is that is part of ongoing activities, such as ongoing work activities. It is frequently used information that is located in the immediate physical or virtual environment.
  - Archived information – is information that is not part of ongoing activities and has been filed away for possible future reference and use. It enables going back and retrieving information for instance from prior activities. It is the least frequently used information.
- Frequency of use of the various classes of information varies widely

The issues mentioned above map out the complexity of personal information management. These issues must be taken into consideration when designing systems, which aim to support and facilitate personal information management.

	Handling and organization	Use	Frequency	Lifetime
Ephemeral	Unstructured Semi-structured	Immediately	Once or few times	Short (near future)
Working	Semi-structured Structured	Active tasks	Frequent	Medium (days to weeks)
Archived	Structured	Later use (not active)	Infrequent	Long (months+)

Table 2.2: Overview of different information classes and a characterization of their properties

Table 2.2 provides an overview of the different information classes and the relations to observed differences in the way personal information is handled,

organized, and used. Also there is a difference in frequency of use and the lifetime of the information. The indications of lifetime are provided as examples.

As the personal information types, structures, and use varies widely, supporting and facilitating personal information management is not necessarily done efficiently by a single organization style. For instance hierarchical structured systems will support archived formally structured information, but does not support unstructured or semi-structured information well.

The following chapter discusses human aspects of personal information management. This includes cognitive aspects, such as memory, recognition and recall. These issues are also relevant to consider in personal information management systems.

Chapter 3

Human Issues

This chapter discusses human issues, that is, human cognitive capabilities. The focus is on issues relevant to personal information management. That is memory, dimensions of memory, recollection, recall, and recognition. The intention is to emphasize cognitive issues, which are relevant to consider in terms of supporting personal information management in interactive systems. This means an emphasis on human issues in relation to tools, which support personal information management.

Although humans may improve their skills by learning and practice, some fundamental limits of human cognitive capabilities exist. Such fundamental cognitive capabilities include short-term and long-term memory and recollection skills. Thus tools designed to facilitate personal information management must take such human capabilities and cognitive issues into consideration.

3.1 Memory

In terms of personal information management *human memory* is the most important cognitive capability. Memory is relevant at different levels and both short-term and long-term memory is relevant to consider.

Short-term memory serves as a way of temporary keeping information. It has very limited capacity and short life time [Mil56]. Due to the short life time

tasks dependant of information held in short-term memory must have limited duration. Chunking information may increase the short-term memory a little, e.g. remembering a phone number is possible, but a similar big number may be difficult.

On the other hand long-term memory holds information for long periods of time and has much higher capacity. Long-term memory manifests itself in terms of human capability of *recalling* and *recognizing* information. However, retrieval from human memory is not addressed by location as is physical filing systems. Rather information is addressed by meaning, that is, the information stored is *content-addressable* or associative [Lan88], which means that information may be recalled based on different elements of the information or associations. For instance humans remember only the substantial parts of events and not the details. In general human memory supports the recollection of the *the meaning* not the *exact details*. Also using mnemonics has good utilization of memory capabilities. The recognition capabilities indicate that human memory store more information about events than we are able to immediately recall [Lan88].

This means that humans may not be able to recall an entire document from memory, but remember the topic of the document, when it was written, and maybe some specific information available in the document. This memory characteristic is an essential reason for the usage of various tools to facilitate the processes of keeping and managing personal information. The limitations of human memory cause the need for an external representation and repository in one form or another of even small amounts of personal information, which can be used on the basis of human memory.

In personal information management the human ability to recall information from memory is therefore highly relevant. Estimates of human long-term memory information capacity have been done [Lan86a]. However, in terms of personal information management the capacity as such is not the primary issue, but rather the ability and the way in which recall and recognition of information takes place. This has implications for facilitating the processes of personal information management.

3.1.1 Known Information and Recall

As mentioned in Section 2.5.2 the purpose of keeping information may be to avoid remembering and repeating a sequence of steps required to obtain the information again. Such a sequence of steps might be long and tedious. Therefore it might be difficult to keep the sequence in memory and recall it accurately for the purpose of later retrieval of the information. Human memory makes such a strategy for keeping information inefficient and unreliable. Instead an approach is the storage of information in some external representation from which the information can be retrieved later.

As mentioned in the previous chapter personal information is information that is handled and kept by an individual with the purpose of enabling such future use of the information. As the information has been handled by the individual and filed by means of individual strategies, it means that the ability to use it later rely considerable on the ability to *remember* about the information and the strategy used to file it. Thereby recalling the original sequence of steps to retrieve the information is avoided. This characterization of personal information allows personal information to be considered *known information.*

The information has some kind of external representation, and the information may be organized by means of an information scheme. However, independently of the external representation and the scheme used for possible organization of the information the recall of information depends on human memory. The individual may have at least partial memory of the information and may be able to recollect something from memory about the information. As mentioned above the human memory of the information is characterized by being associative. This means that retrieval of information from an external representation rely on the ability to recall elements of the information or the meaning of the information. Also the context of the information is important for recall. The context includes the information associations that exist when the information is stored or used [Lan88]. Naturally knowledge about the personal strategy used for storing the information in an external representation is essential too.

3.1.2 Mental Representation

The approach used for filing the personal information therefore assume that the individual can remember something about the information and may be able to recollect something from memory about the information in order to retrieve and use the information. This recollection can include different dimensions of memory about information regardless of the approach used to organize the information. Even if the information was not organized as such using a particular strategy. For instance the individual may partly remember the content of the information, or the time the information was used or received, the person from who the information originated, or a category under which the information was filed.

This means that many different strategies may be used to keep information and information may have many different formats and different medias (paper documents, electronic documents). However, independently of these parameters the personal information kept by an individual has a representation in the memory of the individual. This means that humans form a mental map of the personal information in the external representation and the structure of that information, see Figure 3.1. This includes the relations between the information as perceived by the human. It also includes a representation of the strategy used to store and retrieve information in the external representation, e.g. repository [Col82].

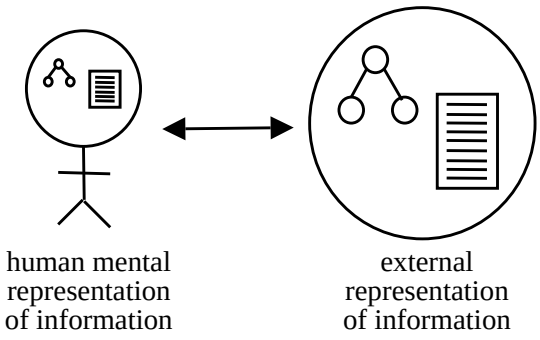

Figure 3.1: External representation of personal information and human mental representation of that known information

In order to keep attention on the task at hand an external representation

or repository must require as little storage and retrieval effort as possible. As mentioned in Section 2.8.3 this is known as the storage retrieval dilemma. As pointed out by Cole [Col82] information must be available at a level so that it allows the development of "a cognitive model adequate for subsequent interactions". The model forms the basis on which storing, finding, and retrieval of information may take place.

The processes of finding and retrieving information are based on context and associations of information available in the mental representation of the personal information. This means that retrieval is based on different types of cues associated with the information.

3.2 Dimensions of Memory

As such humans remember about their information along many different dimensions. Such dimensions of memory include:

- Spatial – where is it located, which place?
- Temporal – when was it used or stored?
- Identification such as a name or a number – what was it called?
- Classification – associated category – how was it filed or categorized?
- Content – what does it contain? What is it about?

These different dimensions of memory may be used in recollecting the information [Lan86b] [Lan88]. The spatial location of an information item is one dimension of the human memory. For instance an individual may recall that an information item was left on the upper-left side of a table, in a particular pile, or on a shelf in the office. He may remember when the information item was last used, which represents a temporal dimension of memory about the information item. Additionally he may partially recollect the content of the information item or the size of the information item by some measure. For instance that is was a document about a particular topic and that it contained about ten pages.

Different dimensions of memory is relevant and additionally both recall and recognition may be used in the process of finding and retrieving information, as mentioned in the previous sections. This means that information management becomes a complex process that includes a combination of techniques and memory for storing, finding, and retrieval of information.

3.2.1 Spatial

As mentioned above one dimension of memory is spatial memory, which means that human spatial memory capabilities can be utilized to store and retrieve information. Those cognitive capabilities are highly developed and strong as humans navigate a physical spatial environment.

In the study by Cole [Col82] it was observed how spatial memory was important as a way to compensate for the difficulties of categorical organization of information. Categorization of information may impose a difficult and time-consuming task upon humans. As a consequence leaving information in a spatial structure is perceive as easier and the mental spatial representation of the information enables humans to retrieve the information by using spatial memory cues. It is perceived that little effort is needed to store the information, as information is simply left in a specific location. Preference for spatial organization has also been observed in the apparent preference for piling of information by some subjects [Mal83].

Little effort might be needed to retrieve the information later, as it can be retrieved at the specific location by means of the spatial memory cues. Searching the spatial information structure is similar to that of navigating the physical environment. This includes the use of landmarks, such as, directional cues, paths, borders, and topological relations. A spatial organization of information includes relations among the information as a navigational cue. Such spatial organization allow for a representation of the relationships among information entities as perceived by humans. The relations between information entities may be mapped to a spatial organization of those information entities in terms of proximity. This means that humans can map their personal understanding, perception and use of the information onto a spatial layout, which support the perceived relations among the personal information. Related information is in

the same pile or the same area.

Both recall and recognition is relevant in the process of finding and retrieving information by means of spatial memory. Recall may be necessary in order to initiate retrieval of information. However, a spatial arrangement of information strongly supports the human recognition of information. This includes recognition of positions and landmarks, and scanning of an area, in order to recognize information.

A potential problem is the scalability of information entities, which can be organized efficiently this way [FFG96b]. As the amount of entities increase it may become difficult to remember exact spatial locations, and the retrieval task may instead become a time-consuming manual search task. Several information entities may have to be visited and searched in order to locate the specific entity needed. In the study by Malone [Mal83] it was observed that some subjects organized their information in many piles, and therefore had to go through piles and search for information to find it. So although human spatial capabilities are strong the organization of information by means of spatial organization may have some limitations in terms of scalability.

A positive side effect of organizing information by means of a spatial arrangement is that the information may also serve the important reminding function. Information can be kept within sight in the spatial organization. The information itself will remind of tasks to carry out. Often humans achieve reminding by means of a spatial arrangement of information by leaving information or objects in locations that are either in sight or frequently visited (see Section 2.4).

In each dimension there is also relations to other information entities. In terms of spatial relations this is the relative position of information entities, in terms of temporal relations these are expressed as "before" or "after", and in terms of semantic relations the relations are how the information entities relate to each other in specific semantic terms.

3.2.2 Temporal

Another dimension of human memory is episodic memory. It is the memory of events in serial form and temporal understanding is a very fundamental cognitive capability. With this type of memory humans are able to reconstruct information

about events on the basis of when it occurred in time. Thus humans may utilize episodic memory of events in order to recall information in terms of *when* it was created, found, or used. In this way temporal information in terms of episodic memory may support the retrieval of information.

Human understanding of temporal concepts is closely related to spatial memory and concepts mentioned above. Temporal properties can be understood in terms of the spatial properties and capabilities. Humans use figurative or metaphorical way of expressing temporal information and relationships. For instance metaphors are used to refer to temporal concepts. That is, spatial concepts are used when referring to temporal aspects. For instance you have the whole day *before* you, or Danish time is *ahead* of British time, and it was left *behind*. Therefore the strong spatial capabilities can to an extent be mapped to temporal aspects. Cues relevant for spatial navigation are also relevant to that of temporal navigation.

The relations among temporal information or events can be expressed in terms of spatial relations. For instance they can be expressed as *before* and *after*, simply meaning that one event happened before or after another event. This indicates how spatial navigation can be utilized when navigation temporal bases information.

This cognitive capability is also powerful, but it relies on the human episodic memory, that is, the ability to recall events and information associated to the events. If no or little information can be recalled about when an event took place this approach fall short. An aspect is the temporal accuracy or level of detail of which information may be recalled, that is, at which temporal level the information is recalled. For instance the year, month, week, day, or hour an event took place. Thus the efficiency of temporal organization of information may depend on whether the recognition aspect can be utilized also. For instance scanning for information based on the temporal level of the information that the human is able to recall.

The spatial capabilities mentioned above allowed good utilization of recall and especially recognition. However, temporal capabilities may have less utilization of recognition due to less navigational cues compared to spatial memory.

Also a potential problem is the scalability of information entities, which can be organized efficiently by means of temporal information. Increased amount of

temporal based information may lead to difficulties recalling information at the level of temporal precision needed. As with spatial organization retrieval may lead to a time-consuming manual search task.

3.2.3 Semantic

Finally human memory includes semantic memory. This type of memory includes facts, concepts and skills, which have been acquired based on experience. As such information has been derived from experience it can be used independently of events (episodic memory).

Humans may associate semantic information to information in memory. For instance create a semantic identification, name or description of information. This may be naming of information entities, creating of identification number, color-coding, and other semantic coding of the information.

Strictly speaking spatial and temporal dimensions can be considered semantic information as well, as they describe "where" the information is or "when" information related to an event took place. However, the spatial and temporal dimension are very fundamental dimensions of information as they map to fundamental cognitive capabilities and understanding. The semantic information mentioned above provides an abstract or logical description of the information, rather than the "where" and "when" of the information.

Information structures and schemes in which information is grouping or categorization is also characterized as semantic information. As discussed in the previous chapter there are many possible ways of doing such classification and many different dimensions on which such classification may be based. In a case study Kwasnik [Kwa89] of how people classify information in an office environment it was found that the *use* of documents was the dominant property on which classification was based. This indicates that situation and context have an important effect on how information is classified rather than the information itself.

3.2.4 Partial Memory

Based on the above mentioned dimensions of memory retrieval of information from long-term memory may take place. This means that humans may be able to retrieve information based on different elements and structures of that information. Additionally the retrieval of information may be based on either recall or recognition or a combination of both.

For instance memory of a particular document could manifest itself in terms of memory of:

- Where it was left (spatial)
- When it was created (temporal)
- When it was last used (temporal)
- The topic of the document (semantic)
- A category under which it was filed (semantic)
- Parts of the content of the document (semantic)

However, studies of personal information management have indicated that humans may only have partial memory of their personal information. This means that although a particular information entity can be associated to many different memory dimensions humans may only be able to partially recall these [Col82].

Lansdale [Lan86b] did exploratory experiments with recalling capabilities based on associations of icons, colours, and locations of a document versus verbal names of category, colour, and name adjectives. It was found that subjects performed equally. The degree of semantic link between the attributes associated and the adverts were important. It was found that icon shapes and category words gave the best recall. Colour and colour names and adjectives and locations were less effective in his experiments.

An important result obtained from the experiment was that these dimensions are *not correlated*. This means that the recall of different dimensions of coding turned out to be *independent*. The likelihood of recalling a dimension of coding

was independent of another dimension of coding already recalled. For example the likelihood of recalling colour was independent of whether the subject had already recalled a shape or location [Lan86b]. In the example above this would imply i.e. that recalling when the document was last used is independent of the ability to recall under which category it was filed. As such human memory of personal information can be characterized as being partial and the different dimensions of memory of the information is independent.

3.3 Storage-Retrieval Trade-offs

As humans may only be able to partially recall information it has implications for the actual practices chosen to keep and file information as discussed in the previous chapter. However, the cognitive issues may provide additional understanding of such different practices of personal information management observed in studies of these.

One aspect is the reluctance to file information away. For instance it has been observed how classification of information is avoided. It is difficult to decide how to actually file or categorize the information. Additionally there is a risk that the information will be irretrievable later. A coping strategy is to put the information in a physical or virtual place. This enables the utilization of spatial memory capabilities when retrieving the information. In physical space also semantic properties may be used as well, such as what did the information look like, e.g. the size or colour of a document. This also allows the utilization of the strong recognition capabilities. This may further facilitate the retrieval process. Additionally the storage of information in a spatial arrangement may be easier than storage by means of a classification scheme in terms of the cognitive burden. If the information is stored into piles this often lead to an implicit temporal order of those piles, as recent information is on the top of the pile. This enables the temporal dimension of memory to be used in retrieval of information. These aspects can explain the preferences for such spatial and temporal arrangement observed in studies of information management behaviour [Lan88].

By using such information management strategy humans do not have to rely on their ability to recall an abstract or logical structure or classification scheme from memory. However, a potential problem of the approach is that it may not

be scalable to a large quantity of information, as the retrieval procedure may become inefficient. For instance much scanning to find the information needed is time-consuming and difficult. But for a small quantity of information the approach has been observed to be useful.

Classification of information is also problematic in terms of cognitive issues. A classification scheme has to be created and maintained, which in itself is a difficult (and time consuming) task [Lan88]. This means that humans have to remember and be able to recall the classification scheme used. The storage of information itself takes away the focus and attention from the actual task at hand. Although that is a general problem of information management it is extensive in storage by means of classification.

As mentioned in Section 2.9.2 limited knowledge of a topic makes it a difficult task to create a scheme that will be useful over time. As the knowledge of the topic expands over time a difficult reorganization of the information may be needed. Or there might be other reasons for altering a classification scheme over time, such as the context of use of the information or other external influence. Even if a topic is well known classification is still difficult for other reasons. It can be difficult to remember and recall labels (properties) assigned to the information for instance assigned names or identification of the information. Barreau and Nardi reported that subjects in their studies were unable to remember filenames of files created earlier the same day [BN95].

Additionally information is not easily organized into neat categorization structures. Information often falls into overlapping categories with no obvious separation. Malone [Mal83] argues that the problem of overlapping categories is what causes little motivation to do classification, as it may lead to problems retrieving the information. He argues that this is the primary motivation for organizing information into piles in the office environment. Thereby there is no need to consider different categories.

The problem of classification is increased by ambiguities in language. Ambiguities can be considered a cognitive problem, as they are partly due to the structure of language, as language contains ambiguities. This implies that category names (labels) can be ambiguous. Thus both storage and retrieval of information may be difficult, for instance if retrieval by categories different from the ones used when filing information. Language ambiguities therefore lead to

an additional cognitive burden, as fairly accurate memory of the meaning of categories chosen must be remembered for the entire classification scheme. Optimally the classification would need to be unambiguous to obtain the highest level of efficiency. Finally the context of the information has an influence on the way it is categorized. The use of information in one context may lead to one classification, whereas another context of use may lead to an entirely different classification [Kwa89]. The ability to retrieve the information therefore depends on the ability to establish a similar context and interpretation of the information for retrieval. As such it depends on the ability to recall both context and interpretation of the information

Storage and retrieval of information is therefore a trade-off among different cognitive aspects and between different difficulties of managing information.

3.4 Retrieval by Recall and Recognition

Retrieval is based on the ability to recall information from memory along different dimensions. As discussed the memory of information is partial, but even with some memory of information this may not be sufficient for retrieval to succeed in all situations. Additional cues may be needed in order to support the human cognitive capabilities and enable retrieval of information.

3.4.1 Interplay Between Recall and Recognition

As discussed above different strategies may be used for filing and finding information: location-based (spatial or temporal) and logical (semantic). In the first the user tries finding the information by recalling its location. The user then navigates to that location for the information. In the study of use of file/document management by Barreau and Nardi it is reported that subjects "often view files in the target location by date, by name, or by some other characteristic in order to identify the file from the list they are scanning" [BN95]. This means that the process of retrieving the information needed is a two-phase process where the initial step is recalling from memory about the information. This may find more information than the particular information needed. Thus the second step is to go through the information to find the particular information needed.

In the example above the initial step is to *recall* and navigate to the location in which the information can be found. The second step is to *recognize* the information among the information available in that location. In the example sorting the available information by different attributes serves that purpose. Such retrieval of information utilizes combinations of recall and recognition. Initially recall is used. Whereas recognition is used in order to find the particular information entity from a list. However, the second step may also include further recall. In the example recall of approximate date, name, or other characteristic may help the process of recognition from the list.

Semantic coding of information may not be the one best suitable. As mentioned earlier it was also found in the study [BN95] that often subjects were not able to recall file names even for files created the same day. Recognizing a file name from a list was preferred instead. This is in accordance with the finding that recognizing requires less cognitive effort than recall [EK95].

Thus recall and recognition are both aspects of the human memory structure. Recalling is active in the sense that it requires searching for information and retrieving it from memory. Contrary recognition is passive in the sense that a piece of information is compared to information stored in memory. Recognition is less demanding from a cognitive point of view, and is more powerful than recall [EK95]. For instance when humans try to recall a set of names from memory they might not be able to recall the whole set, whereas if they are provided with a list of the names, they will recognize a larger set of names. This indicates that naming of information entities is not always an efficient way of organizing and storing information. Some entities may not have a natural name, and coming up with one is artificial and thus easily forgotten, as found in the studies by Barreau and Nardi [BN95]. Although personal information is considered known information there may be situations where humans may not be able to recall sufficient information, which enable actual retrieval of a specific information entity. In that case a piece of information can be said to be *forgotten*. However, the term is unsuitable, as providing additional information cues may actually enable humans to remember the information later. Thus additional means may be needed to enable retrieval.

3.4.2 Reminding

In the previous chapter it was discussed that reminding is an important aspect of personal information management [Col82] [Mal83]. In the discussions above retrieval has been discussed as an active process initiated by humans. Reminding can be seen as an information retrieval process as well. In that case it is a passive process, which is not initiated by humans.

The observations of recall and recognition as memory aspects are important for the reminding aspect of personal information management. In order to be reminded of information humans naturally rely on recognition rather than recall, as recognition is the passive process and recall is the active process. For instance reminding allows avoiding actively remembering information, as the good recognition capabilities can help serve reminding.

As mentioned in the previous chapter studies of personal information management has indicated some preference for spatial organization of information in order to support reminding. As discussed above spatial organization strongly utilizes the recognition aspects of memory. That is, when navigating to a particular spatial location it enables related information to be recognized and used. This also serves the reminding function of personal information management, in that humans are reminded of information when recognizing information in a particular context or situation of use.

3.5 Classes of Information

Finally the different classes of personal information, ephemeral, working, and archived information discussed in Section 2.7 are revisited and discussed briefly from a cognitive point of view. The different classes of personal information also imply different levels of interaction, that is, different handling and use of the information. This means information in different situations of use, and over different time-spans, which means it has implications for the utilization of different aspects of human memory.

3.5.1 Ephemeral Information

As discussed in Section 2.7.1 ephemeral information is information that is mainly used on a short-term basis and it is also referred to as action information. Therefore it may involve little memory as such, as the information is typically in sight and typically used immediately. Therefore retrieval of ephemeral information relies more on recognition that on recall. This is also supported by the observation that it is typically not stored away, but rather handled in an ad hoc manner.

However, dealing with and organizing ephemeral information can be problematic if it is not supported well by the tools used to support personal information management. Therefore observations have shown that it is often done in a very ad hoc manner [BN95]. Therefore the limitation of short-term memory [Mil56] is relevant cognitive issue.

3.5.2 Working Information

Working information is used frequently compared to the other classes of personal information (see Section 2.7.2). This information is typically relevant to ongoing tasks, and usually organized in a more structured way than ephemeral information. This implies that retrieval of such information rely more on recall than ephemeral information. However, spatial organization of the information may be used and thus both recall and recognition are important aspects. Additionally the management and use of this information typically rely on both short-term memory and some degree on long-term memory. As the information is frequently used it is easier to recall the storage locations of information, and a possible classification of the information.

3.5.3 Archived Information

At some point working information turns into archived information (see Section 2.7.3) if it is decided to keep the information for possible future long-term use. However, as the information is infrequently used more systematic organization of the information is typically used. Observations by Cole [Col82] and Malone [Mal83] have shown that typically some kind of classification systems is used to archive the information rather than spatial or temporal strategies for

filing. Therefore this class of information relies on long-term memory and the creation of schemes, which enables the possible later retrieval of the archived information. As discussed in the previous sections this causes a cognitive burden on humans as the creation and maintenance of a classification scheme is difficult.

3.6 Summary

The most important issues in personal information management discussed in this section can be summarized as follows:

- Humans have partial memory about their personal information (along several dimensions), which means that supporting systems should make it possible for humans to utilize this partial memory
- Organizing information is difficult – coming up with a good classification scheme is difficult. Archiving information (using the scheme) is time consuming and difficult.
- It might be ambiguous how or where the information fits into the classification scheme.
- Information and information structures changes over time.
- The context in which information is used change over time.

In the following sections we discuss how different models and computer-based approaches to support personal information management support these issues. This means how different approaches support the cognitive aspects involved in personal information management.

Chapter 4

Systems

As seen in the previous chapters personal information management is a complex area as it involves information at many levels and kept and used under different circumstances. There is a diversity of personal information used in relation to different activities. The level of structuring of the information varies and information is interacted with in different ways. Also the frequency of use varies widely.

When considerable information quantities are dealt with as part of daily activities some kind of external system must be used to support and facilitate the management of personal information. This is due to human cognitive capabilities such as memory capabilities as discussed in the previous chapter. The issues and aspects involved in personal information management have implications for the way in which personal information management can be supported by both paper-based and interactive systems. Supporting and facilitating the processes involved in personal information management are therefore not trivial.

The quantities of personal information and the human cognitive limitations (e.g. memory) are reasons that tools supporting personal information management are needed. The more time spent filing and managing information, the less time is spent performing the actual tasks at hand. A challenge for personal information management systems is therefore to facilitate the processes involved in a way so that the time spent filing and managing information as such is minimized.

Different approaches and systems to support and facilitate personal information management have been tried out both in the research and commercial space over the last decades. This chapter discusses these different general approaches and systems from the perspectives of personal information management and cognitive issues, which have been discussed in the previous chapters. Advantages and disadvantages in terms of supporting the issues involved are discussed.

4.1 Expressing Intent

Systems that aims to support and facilitate personal information management needs to address a number of issues including human cognitive issues. Supporting systems basically provide an external representation of the personal information. At the same time the human has an internal representation of the personal information in the sense that he is able to recall the information along different dimensions. This may include recall of parts of the information, when it was last used, the context in which it was used, or other information associated with an information item. That is, the internal representation in human memory includes relations to context and experiences. These entities are tacit knowledge, which is inside the human memory only, and does not have an external symbolic representation.

Therefore personal information management involves the interplay between the human internal representation of the information and the external representation in some external system. An issue is the gap between how humans perform personal information management and how the system actually support the processes involved in personal information management and how easy it is for humans to map their personal information management intents to interaction with the system. Thus an essential problem that systems need to address is how humans can express their intent. On the basis of the task at hand and the internal representation of information humans express intent. This intent must be mapped onto the system and the external representation. The essential issue is to bridge the gap between what humans know and can recollect about their personal information from memory and how they can map that knowledge onto interactions with an interactive system, in order to retrieve, manage, and use information in an external representation. Figure 4.1 illustrates processes

involved in personal information management, including the interplay between the human (internal representation of information) and the system (the external representation).

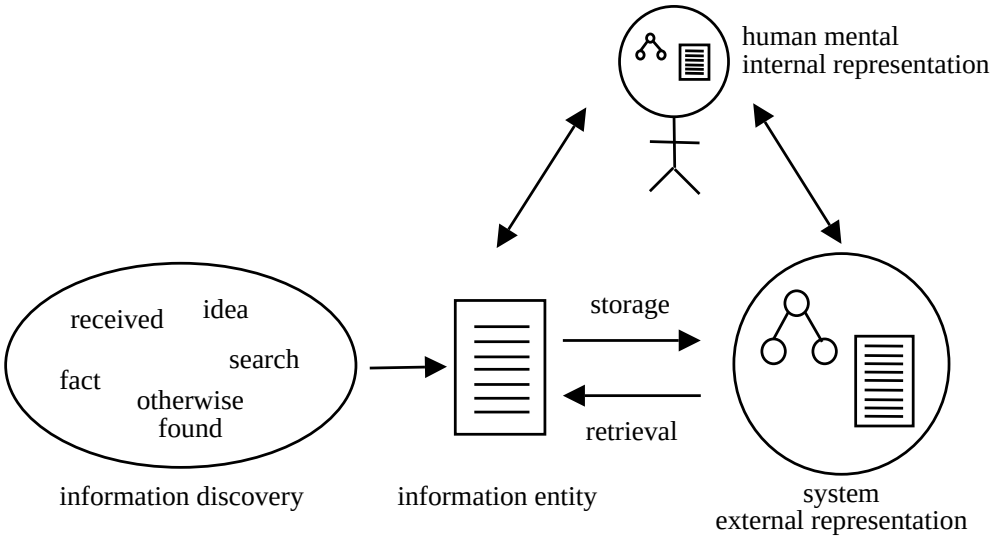

Figure 4.1: The processes involved in personal information management and the mental representation of known items

The processes include information discovery, information item, storage and retrieval of those items in an external representation, and the human mental representation of those know information items. The expression of intent is the mapping between human and system.

Systems that aim to support personal information management must therefore utilize and support this interplay and human cognitive capabilities in order to bridge this gap. The differences in the two representations mean that humans perform a mapping of expressing intentions from the task domain (source) to the actual interactions in a system domain (target). Humans typically remember far more about their personal information (along many dimensions) than they are able to use in retrieval procedures in interactive systems [Lan88]. As

an example some approaches only allow humans one or few ways of expressing how to retrieve information from an external representation. Similar storage of information is only allowed along one or few dimensions in many approaches.

Humans may be capable of expressing many features about the information that for instance interactive system cannot use as input in order to retrieve the information. This implies a cognitive load on the human to convert that knowledge about the information into something that can be used as input in the system. That means converting the intent into an expression that the system understands.

Example: "I know about the project that it has information about X, Y, and Z. But I don't remember when it was used or under which category it was filled. So now I need to search the entire system for X, Y, and Z."

Example: "I had the phone number for this person, but is it stored in an e-mail message, in a document, in my address book, a bookmark to a webpage that has it, or in my calendar?"

The mapping is influenced by the constraint of the possible ways of interaction with the system. The further the gap is and the complexity of the mapping the more it requires the user to take attention from the task at hand to the actual specific interactions of the system. One issue is the specific way in which the external representation is used. Another issue is that interactive systems may have specific terminology, such as, commands, files, applications, cursor, etc. The user is familiar with a source domain, and the terminology used in that source domain.

The aim of systems supporting and facilitating personal information management is therefore to bridge the gap and reduce the complexity of the mapping of intent in the task domain to the operations on the tools, thus leading to a more efficient and usable personal information management system.

4.2 Supporting Personal Information Management

Personal information management has been studied for the last couple of decades. Initial studies of such as those by Malone [Mal83] and Cole [Col82] involved observations of office work at a desk and paper-based systems for the support of

information management. More resent studies have involved personal information management of electronic information, such as, management of emails and bookmarks. Therefore the information environment has changed considerable over the last decades, as more information has become available through computer networks.

However, the fundamental issues involved in personal information management are essentially the same whether the information is managed in a paper-based system or in an interactive system. Therefore it is important that both types of systems address the issues involved in personal information management and human cognitive capabilities discussed in the previous chapters in order to provide usable support.

In the interactive systems domain the types of approaches to personal information management can be divided into two groups:

- The first includes systems that are characterized by an approach that is similar to a physical personal information tool or systems based on a metaphorical approach
- The second includes systems that are not based on a physical tool, approach or directly on a metaphor

In the first group belong a group of personal information management specific software and the traditional desktop system approach [JRV+89]. In the second group belong approaches such as Lifestreams [FFG96b], Placeless Documents [DELS99], Haystack project [AKA99] and Lotus Agenda [KKB+90]. Together these approaches represent a fairly wide spectrum of different approaches to personal information management and together they highlight important issues involved in designing such interactive systems.

4.2.1 Personal Information Management Specific Software

Traditional types of personal information would be kept in a paper-based form in terms of information in address books, calendars, to-do lists, and diaries. In a similar way software applications have been created to support personal information management. Initial interactive systems have tried to mimic the traditional approaches and tools. Some of these systems have been designed to closely resemble physical paper-based systems, such as paper organizers.

Such systems can be characterized as tailor-made systems, as they serve the purpose of supporting very specific personal information management tasks and functions. They map functions from the traditional paper-domain directly to functions in the interactive systems domain. To the extent possible the same set of functions are available in both domains. Additionally the functionality is divided into the same areas and sets of functions with clear boundaries among functionality. For instance separate areas that deal with addresses, calendar, and to-do items.

An example of this approach is computer desktop application software designed to be very similar to traditional paper-based systems. For instance desktop software systems, such as Lotus Organizer (see Figure 4.2) and others closely mapped the functionally of paper-based tools into the software approach. Especially visual features where included in order to closely resemble the visual appearance of a paper-based calendar and organizer.

Examples of functional features and visual similarities include:

- Clear division of functional areas (address, calendar, to-do items)
- Visual appearance – book with spiral binding and paper pages
- Pages looking similar to paper-systems (units of information)
- Alphabetic index pages
- Index pages similar to paper-based systems (with different colors)
- Trashcan to discard information

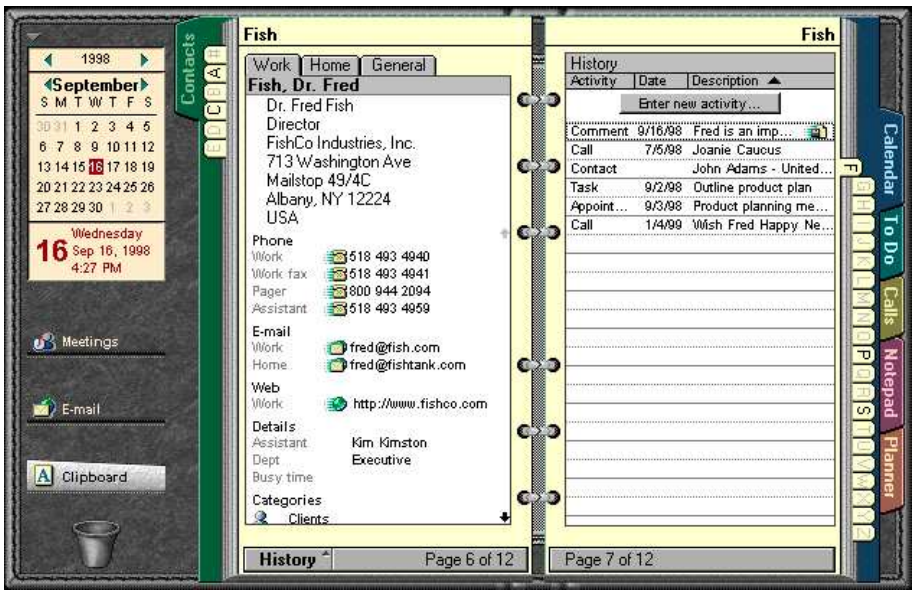

Figure 4.2: Screen shot of the Lotus Organizer software application

Some of these visual features are of course unnecessary parts of the user interface from a strictly functional point of view. The book and spiral binding does not serve any functional purpose, only a visual purpose to underscore the paper metaphor used in the system.

A potential benefit of such a metaphorical approach is that it allows the transfer of knowledge from a source domain (the paper-based systems) to a target domain the software system. Thus the purpose of this strategy is to allow the users of systems to be able to quickly learn to use the system and user interface. If users are used to a paper-based organizer system they should be able to easily recognize the corresponding elements in the software system and use those elements. The software system has functionally that is similar to that of a paper-based system. It might have additional functions such as backup and search mechanisms and the ability to show calendar information in different month–, week–, and day-views. However, flicking through the pages of the paper-based system might be faster than searching or shifting pages in

the software system. Also the software only allow the user to input information in specific areas and fields. The user must navigate to those fields and input information, whereas the paper organizer allows the user to write a quick note or annotation anywhere on any page.

The approach clearly has both benefits and disadvantages. However, in a comparison of the paper-based and software based systems issues which influence situations of use must be taken into account:

- Physical factors (size and weight)
- Interaction (pen vs. keyboard)
- Situation of use (stationary vs. mobile tool, places of use, standing, sitting, et cetera)

The software approach might have the drawback of not being portable like the paper-based organizer. Therefore the user is limited to a stationary situation of use typically behind a desk using a computer desktop system. Especially the issue of portability potentially yields two different situations of which are unsuitable for direct comparison. The paper-based solution clearly has advantages in mobile situations of use and easy writing notes and annotations anywhere. Whereas the electronic solution may have benefits in an organization context where employees need to view the calendar of colleagues and simultaneously perform booking of meetings in a number of calendars. It is noted that recently mobile interactive personal information management solutions have become available which makes the comparison of the approaches more complex.

Nevertheless it is relevant to consider the functionality of the approaches in the two different domains. The main benefits and disadvantages can be summarized as follows. The approach discussed here clearly has the metaphorical benefit of allowing transfer of knowledge from the paper domain to the interactive domain. This potentially allows those familiar with paper-based organizers easier learning of the software-based organizer. However, the metaphorical approach is also the disadvantage of the approach. By closely simulating physical artifacts such solutions does not necessarily take advantage of the possibilities in the interactive systems domain. The paper metaphor may introduce a set of

constraints on the interactive system that are unnecessary and prevents utilization of the benefits of the domain. Such limitations could include the navigation, visualization, and use of information in the system. For instance there could be more efficient ways to navigate the information than turning virtual pages, and other ways of visualizing the information than through units of paper. Therefore such an approach to the construction of personal information management systems may be imposed by a set of artificial constraints and not fully utilize the potential of the interactive system domain beyond a few features as mentioned above.

In general these traditional PIM software applications does not necessarily utilize the full potential of a computer-based system. Instead of simulating the physical paper-based systems used to support personal information management other approaches might provide better support for the activities and processes involved.

4.2.2 The Filing Cabinet Metaphor and File Systems

Many different user interfaces for computer systems exist. However, a hierarchical file system is the typical underlying storage mechanism in many computer operating systems. Such storage system is based on the physical filing cabinet metaphor with drawers, folders, and papers.

This approach transfers the (manual) management processes that take place using a physical filing cabinet to a computer based one. Thus it does not necessarily utilize the possibilities in a computer-based system. Classification of information and finding a place (typically only one) to put an information entity is required. This is difficult from a cognitive point of view as discussed in Chapter 3. Also it has been identified in studies of personal information management that "much of the organization on people's desktops consists of untitled piles, bit that most computer systems require any new document or group to be titled" [Mal83]. As has been discussed studies of personal information management have shown that information handling and use includes a variety of information types and levels of interaction. The frequency of use also varies and the information structures and use may change over time.

As such there are a number of reasons why hierarchical file systems may

not support the full range of personal information efficiently. This explains why coping strategies have been observed in these studies as well.

4.2.3 Desktop Approach

For the past decades the desktop paradigm has been the most widespread metaphor for computer user interfaces, that is, the user interface to personal computing operating systems. The desktop paradigm can be considered an approach to personal information management. However, when discussing the desktop approach in terms of personal information management support, it has to be observed that the primary goal of the desktop model (when it was introduced) was not to support personal information management as such, but rather that it should be easy to learn to use the system [JRV+89]. That was one of the ideas behind the metaphor, that is, a way to ease learning to use the user interface. Knowledge from a source domain can be used to in the target domain, that is, the user will have a better chance to guess or understand how the user interface works. That process should be faster because of the knowledge from the source domain. The user of the system would be familiar with concepts from the office domain and be able to transfer the knowledge about those into the target domain.

This explains the metaphorical approach, which tried to take elements from the traditional office environment thus enabling faster, and easier learning of the new computer-based tools. These elements included the filing cabinet, paper documents (files), folders, and a trashcan. An additional set of metaphors is found in the windows, icons, menus, and pointers (WIMP) interface elements. A metaphorical approach might facilitate learning, but a potential problem is that the metaphor eventually breaks.

The paradigm was also introduced in order to facilitate office automation, as it tried to map existing work practices, such as document handling and the filing cabinet in a hierarchical file system in the electronic domain. A goal was to facilitate the process of creating paper documents, why software applications typically are document centered, and enable the creation, handling of paper documents, and printing. Typical examples are so-called productivity software such as spreadsheet and word processing applications.

However, studies have shown several usability problems in the "desktop paradigm" approach. The desktop paradigm does not necessary scale well to the situations of today, where computers are networked, the information quantities are higher, and the available functionality in terms of available commands are orders of magnitudes higher.

In the beginning the desktop model was applied with only a few application programs: word processing, spreadsheets, and presentation software. Each of the applications had limited functionality as they had a relatively small number of commands compared to similar applications today. However, each new release of the application software typically introduces additional functionality and thus many new commands. Thereby making it more difficult to learn and use [Ras00] [Nor98].

In addition the file system held only relatively few files, so the hierarchical file system did not present a major problem in terms of usability initially. As the number of files in the file system is low and the number of available commands in the application menus is low then the model is usable. The reason is that the user can easily "scan" menu item or information needed, since there is a low number and it fits on the screen.

As, applications have increased in functionality by adding additional features and thus the number of available commands in each application has increased. Therefore application menus were changed into hierarchical menus in order to fit the increasing amount of commands and options into the available space on the display.

Although the desktop paradigm was not created to enable personal information management as such, it is actually used to serve personal information management purposes. Studies by Nardi and Barreau [BN95] have shown the different classes of personal information use are found in the use of desktop systems too. Also users use strategies in which desktop systems can serve the reminding function, such as, leaving icons in special places on the desktop [BN95]. Another ad hoc method is to leave e-mails in the inbox to serve the reminding function. As e-mail was originally intended to support asynchronous messaging only, this clearly indicates an example of unintended use [WS96]. Although the system was not designed to enable these personal information management features users of the system have developed coping strategies that enable them to use the system

in unintended ways for the needed purposes.

Today the desktop metaphor and variants of the scheme is the most widespread and common user interface for computer systems and handheld devices such as personal digital assistants. As this approach mimics the physical desktop and office, it requires the user to do filing and management of information manually. Filing is done similar to the way this would be done using a physical filing cabinet. The approach was designed when computer hardware and storage was limited, and therefore only had few software applications and file systems held limited quantities of information.

Current systems have become much more powerful. The number of applications and the amount of storage has increased drastically. Thus desktop computers must handling of orders of magnitude higher quantities of information. Although desktop operating systems have changed over the last decades, the user interfaces are still built upon the same underlying model. Additionally the underlying mechanism for storing information into the system is still based on the filing cabinet metaphor, that is, a hierarchical file system with files and folders.

This approach does not appear to scale well with increasing quantities of information and types of information. Additionally the approach is not tailored for the support of personal information management as such. File systems may provide some level of support for archived information, but ephemeral and working information has less support [BN95]. Similar the support for loosely structured information is not built into the model, why users have to develop coping strategies. Also to enable support of reminding users have to use the system in alternative ways. Finally a strict hierarchical information structure is limited in the way storage and retrieval may take place, and thus does not enable taking full advantage of the different dimensions of memory involved in recall from human memory.

4.2.4 Alternative Approaches Researched

Over the years other approaches to the support of personal information management have been researched. These alternative approaches have explored how to utilize advantages in the electronic domain. This has included research of information retrieval and different ways of organizing, visualizing, and interacting

with information stored in personal information management systems.

On area of research that has received much attention is information retrieval. However, the main concern of this branch of research has not been on retrieval of known personal information from information systems. Rather it has assumed that the information is already in the system – in a database and mostly not entered or stored by the user of the system. In that case the user typically has to learn a classification scheme, or search keywords to retrieve the information.

Entry and handling of information especially of ephemeral information, that is, active non-archived information has gotten less attention in HCI research [BN95]. However, a research area that has gotten some attention is information visualization, for instance approaches that enable visualization of large information quantities for easier information retrieval and exploration [Shn96].

These general research areas mentioned above have some relevance to personal information management, however research that address the particular problems of personal information management have also been carried out over the last decades.

Especially different approaches to personal information management have been researched over the last decades. This include research that have proposed novel user interfaces that aim to improve interaction and provide better support for the specific characteristics and issues involved in personal information management. Thus facilitating the process of information organization, managing increasing quantities of information and varying information types. Additionally supporting the increasing amounts of ephemeral and working information that information workers deal with on a daily basis.

These approaches to personal information management include:

- Temporal-based systems
- Spatial-based systems
- Attribute-based systems

Systems that support personal information management in general typically support multiple information types in a combined system, for instance a com-

bined calendar, address book, to-do, and notes system, as discussed in Section 4.2.1. Other systems support only a specific domain of personal information management. An e-mail system that supports the management of e-mails and collections of e-mails is an example.

The approaches discussed below address the problems of personal information management and therefore typically take into account the relations among different types of information, and that information is related across activities. Therefore these approaches typically address the management of an entire domain of personal information, including information of different type rather than addressing the management of a single information type only.

The overall purpose of these systems is to support and facilitate the processes of information management. Having the user to focus less on the information management as such, and more on the information use, and the actual task at hand.

4.2.5 Temporal

One area that has been researched is to utilize different dimensions of memory. That is to allow the users to organize information in different information structures, and not necessarily enforce a hierarchical information organization and explicit naming of information entities.

One example is the temporally based approaches to personal information management. Examples of such approached that have been researched are:

- Lifesteams – address the problem of having to name objects, categorize them, and store them into a hierarchical structure [FFG96b].
- Time-machine computing from Sony research [Rek99] – the starting point is the traditional desktop metaphor. The extension is to allow the user to move backwards in time and retrieve information from the desktop in an earlier context
- Pile metaphor – an approach from Apple which support unstructured and semi-structured information in terms of piles [MSW92]. These piles are

similar to streams in terms of being implicitly temporally organized information (but not necessarily strictly so).

An alternative to the desktop model is the Lifestreams model [FFG96b] that addresses the problems introduced by hierarchical file systems. The Lifestreams model was originally a research project from Yale University, and later commercialized by Scopeware (previously Mirrorworlds). Lifestreams is described as a network accessible data archive for time-ordered information – a time ordered stream of documents. A suitable metaphor for describing the system is an "electronic diary" metaphor, meaning that the overall organization principle is the temporal dimension.

The unit of work in Lifestreams is documents. Document formats include those found in a regular computer desktop environment, such as, word processing documents, spreadsheets, reports. However, it also includes information such as memos, calendar entries, e-mails, et cetera.

The goal of the system is to decrease or eliminate the time spend managing information, thus enabling the user to keep attention to the task instead of the information management as such. Using a temporal organization of user documents in the user interface the user is no longer required to explicitly name files or organize files into a hierarchical structure. Instead all files are organized into a linear stream by temporal attributes, such as, the time when the file was created or used. This means when a new document is created it is not necessary to provide a filename or put it in a specific location, as is the case in the desktop model. Documents are automatically archived by the temporal property. Thus some level of ephemeral and working is available as such information is in the "present" front of the stream and highly visible to the user, while older information is automatically pushed into the past further from the user's view [FFG96b].

Retrieval of documents in the system is done based on temporal attributes. The design rationale is that documents are easier to find if retrieved in the same (temporal) context they were created in. The argument is that in retrieval from a document hierarchy it might be ambiguous where the document is located, as discussed in Section 3.3. In addition the context in which the document is used might change over time. These are among the reasons for the time-based notion in Lifestreams.

As discussed in Section 2.4 an important aspect of personal information management is reminding [Mal83] [BN95]. In Lifestreams it is argued that reminding is an integrated part of the model, as reminders are simply put on the stream in the future, and will automatically appear "on top" of the stream, that is, in the front of the current stream. This makes it visible to the user. These so-called "future documents" can also serve as placeholders for meeting schedules [FFG96b]. This approach is suitable for information or a reminder where a temporal property is naturally associated, but seems less suitable for information or reminders where no logical temporal property can be associated.

Time-based user interfaces were suggested as a design possibility from the findings in the study by Malone [Mal83] mentioned earlier. It was suggested that the temporal parameter is used for information classification. Either based on the time the information was created or by the time it was last used. "One can imagine a system where users search for a document by kind of simulated time-lapse photography of the history of their desktop. They could 'rewind' and 'fast forward' the desktop to locate the last time the desired document was on the desk" [Mal83]. The Lifestreams approach resembles the proposal by Malone.

However, a system that is even closer is a *time-centric workspace* found in "Time-machine computing" from Sony research [Rek99]. The potential of finding information from an earlier temporal context is researched. The starting point of the information environment is the desktop user interface. Then the system allows the user to move backward in time (similar to Lifestreams) to get a view of the desktop information environment in an earlier state and context. The idea is to support retrieval by establishing an earlier context in which the information was used.

Finally the "pile metaphor" project from Apple computer is also relevant from a temporal approach perspective. This user interface model is also based on a diary metaphor [MSW92]. As discussed in Section 2.6.2 piles may have an implicit temporal order. Therefore this approach can be considered time-based as well.

Lifestreams and Time-machine computing seem to have a strong emphasis on supporting retrieval of archived information. As the time-machine computing approach is strongly based on the traditional desktop approach it seems it would have some limitations in support for ephemeral and working information.

Lifestreams support reminders in terms of information appearing at the front of streams and substreams. However, in general the approach appears to have only some level of support for ephemeral and working information. The Lifestreams approach has advantages over the traditional storage in hierarchical based systems as it removes the requirement for explicit naming of information and explicit organization of information in a hierarchy. Also it allows information of different types relating to the same activity to be stored together in streams. A potential drawback of the solution is that the focus is on temporal based retrieval, thus utilizing only the temporal dimension of retrieval. This requires the user of the system to be able to recall along one dimension, that is, when information was used. It is noted that some search facilities are available in implementations of the Lifestreams approach, thus enabling alternative ways of retrieving information from the system via generated substreams.

The pile metaphor appears aimed as an addition to existing systems, thereby enhancing the support of unstructured and semi-structured information in existing systems such as desktop systems. As Lifestreams it avoids forcing users to explicitly name information entities, but rather keep them in a semi-structured sequence – the pile. Thus the pile metaphor may enhance the level of support of unstructured and semi-structure information in existing systems. These are important properties of personal information management.

4.2.6 Spatial

Another dimension of human memory that can be utilized is spatial memory. In this approach information is stored in a spatial way in a two– or three-dimensional virtual environment. Retrieval of the information is based on the human capabilities to recall spatially where information is located.

Examples of systems that use this approach includes Zoomable interface PAD++ [BH94] and Data Mountain [RCL+98]. Both represent systems that enable users to organize documents in a spatial virtual environment. PAD++ provides a zoomable surface onto which documents can be placed. The user can pan the zoomable areas and zoom in and out on the surface in order to view documents in larger detail. One application of the PAD++ zoomable interface technique is for personal information management. In that case the personal

information is managed by placing it onto the two-dimensional surface. Relations of information items is achieved by spatial proximity, for instance storing related information items in a chosen area on the two-dimensional surface. Data Mountain is a three-dimensional document management system. It enables organization of bookmarks (webpage documents) in a spatial environment, but the system is general in the sense that it allow the organization of other types of documents as well.

An advantage to these approaches is that it enables a flexible way of storing information in different levels of structure. It does not enforce a hierarchical information organization and explicit naming of information entities as information is simply organized spatially. However, the approach does enable the creation of logical groups of related information in terms of spatial proximity. If a hierarchical structure is needed it is possible to map this to a spatial information structure for instance in the PAD++ approach. The surface can be divided into any number of areas where information is located together. Levels in the hierarchical structure can be achieved by zooming in on a particular area that again can be divided into any number of sub-areas. Thus the structure of information in the PAD++ approach can be said to have an implicit hierarchical structure. At the same time the approach allow less structured information as well. As such the approach supports information that have varying degrees of structure, as information can be structured, semi-structured, and unstructured. As the approach enables less structured information is may also be more suitable for ephemeral and working information.

As a final example of the spatial approach the desktop approach discussed previously can be mentioned. This also involves a spatial approach in the sense that it utilizes some elements of spatial organization. For instance placing documents or icons on the desktop itself represents a two-dimensional spatial environment. As have been seen in studies this is typically used to enable reminding. Thus the spatial approach is advantageous in terms of reminding as studies have pointed out the preference placing information spatially in order to be reminded [Mal83] [Col82] [BN95].

Utilizing human spatial memory capabilities seems to be an attractive approach, as those capabilities may support information management well, and also resemble how information is managed in the physical environment. Also

human spatial memory capabilities and visual perception are strong (see Section 3.2.1).

One practical problem that spatial based systems have had to deal with is that implementations need powerful computers to render a spatial environment. With less powerful machines this can potentially cause long response times when navigating the spatial environment, which then becomes a cumbersome and slow process. Thereby navigating and using the information appears inefficient. As computers become more powerful this specific issue will be less of a problem and the spatial navigation may support reasonable response times.

With zooming interfaces all information fits onto one display area, but with little detail at the top level. In order to get the details of the information the user has to zoom in to bring specific information into focus and detailed view. Thus information is shown with different levels of details depending on the depth of view. The process of navigating a virtual spatial environment involves moving around the spatial structure by panning and zooming, which may cause the processes of information storage and retrieval to take longer compared to other approaches. Another issue is that it may lead the user to the feeling of getting lost in the information environment. Implementations of zooming interfaces typically include a feature that allows the user to return to a home position or known starting point.

It is questionable if utilizing only spatial capabilities may provide support for the full range of information and issues involved in personal information management. Additionally the scalability of the approach in terms of information quantity is an issue. For large quantities of information other means of information organization and search facilities may prove to be more efficient. An effective search mechanism may be more efficient than navigating the spatial environment in the ways described above. Further studies of use of such systems for personal information management would be required in order to establish the usability over long time use and for considerably quantities of information.

However, that does not mean that the spatial approach is inefficient for all situations involved in -personal information management. Especially in terms of supporting reminding in personal information management tools spatial organization of information seems to be highly relevant. Also including some level of spatial organization of information may enhance support of ephemeral and

working information. Avoiding naming and classification of such systems is an important issue (see Section 3.4.1), and the spatial approach is one way to achieve this.

4.2.7 Attribute-Oriented Approach

A critique of the approaches discussed above is that they are limited in the way in which they allow storage and retrieval of information. The approaches only utilize few dimensions of human memory. If the user of a system is only capable of recalling information partially it may be difficult to perform retrieval in the system. For instance if the user only recalls when information was used retrieval is difficult if the information is stored in a spatial-oriented system. This may lead the user to a time-consuming manual search of the virtual environment in order to retrieve the information needed. Search facilities could be available in actual system implementations. Since the user has to search for the information by other means it could be an indication that the approach is not working or that the approach is not suitable for the type of information stored in the system.

Another problem is that information can only reside in one place. For instance a document can only be placed in one particular location in the spatial environment. However, the information may belong naturally in more than one context, and thus be organized in multiple ways as part of different groups of related information. Finally enforcing a hierarchical information organization may not be suitable for all types of personal information. This approach is also limited to information in one location and thus one way of finding and retrieving the information. In file systems files typically have a set of standard attributes such as the type of information, the size of the information (e.g. file size), the date of the creation of the information, the date of the last modification of the information, name of the person who created the information, et cetera. These are properties that can be used in the retrieval process once a group in which the information resides (see Section 3.4.1).

The issues mentioned above have been addressed in another branch of approaches to personal information management and document management – the so-called attribute-oriented or associative systems. In these systems information is organized by associating descriptive attributes to the information entities.

An example is the semantic file system [GJSJ91], which is an extension of a traditional hierarchical file system. In this approach attributes are automatically extracting from the content of the files in the system. The attributes are indexed and enable the user to establish queries on the file system, and thereby retrieving files from the semantic file system by means of the keywords associated with the files. Thereby a number of *virtual directories* can be established, that is, a more flexible way of retrieving information from the system compared to traditional file systems. This means that the information may be retrieved in multiple ways, as the same information may be available in multiple virtual directories.

The semantic file system represents an example of general file management and is not specifically meant to support personal information management and the specifics of such systems. There are other similar approaches that are targeted towards supporting personal information management. These approaches include:

- Lotus Agenda – Personal information management software from Lotus [KKB+90]
- Placeless documents – Research project from Xerox [DELS99], [DEL+00]
- Haystack – Research project from MIT [HKQ02]
- Lifestreams – can also be categorized as an attribute-oriented approach [FFG96b]

Although these systems generally provide support of personal information management they are different approaches that each emphasize different sets of aspects of personal information management. Nevertheless they share some similarities in terms of the underlying storage mechanism and means of retrieving of information. Common among the approaches is that they use some kind of attribute-oriented scheme. This generally means some way of associating multiple descriptive attributes to information entities kept in the systems. This typically enables a more flexible way of organizing and retrieving information compared to traditional hierarchical-based approaches. For instance information may be organized in multiple ways based on the attributes associated with the information entities. This enables information to appear in multiple contexts

of use. The approach therefore potentially provides a better support for retrieval of information, as multiple dimensions recalled from human memory about the information can be used in order to perform information retrieval from the system.

Lotus Agenda [KKB+90] is an example of personal information management specific software. Although this software is DOS-based with a simple text-based user interface the approach in this software is somewhat different from the typical examples of personal information management specific software mentioned in Section 4.2.1. The aim of this approach is to a larger extent to address some of the specific issues involved in personal information management. The system handles simple traditional personal information such as calendar items, to-do items, memos, addresses, tasks, calls, et cetera. However, it is different as it uses an attribute-oriented scheme referred to as an item/category database. Where traditional systems have separate system-defined areas and views of each type of information such as addresses and calendar information, Agenda have no fixed boundaries or areas based on information types. This means that related information of different types can be organized together in so-called views, which can be created and tailor-made by the user of the system. Thus Agenda can be characterized as having a much higher degree of flexibility in terms of structuring, organizing, and using the information kept. This also means support of less structured information, as the information entities does not have to be explicitly named or organized into a specific structure. Additionally the approach allows the information structures and contexts in which information is used to evolve over time. Finally Agenda appears to have a higher level of support for ephemeral information, as it allows easy and quick writing of information such as a phone number or short piece of text into the system.

Placeless Documents from Xerox Xerox [DELS99], [DEL+00] addresses some of the problems in hierarchical-based systems. Specifically it addresses the problem of having to place information into one specific *place* in a hierarchical information structure (or other information structure), and the requirement to afterwards retrieve it by navigating to that one specific location in the structure. This includes the issue that location of information may be ambiguous as information may belong in multiple locations. The aim of the approach has lead to the name *placeless* documents, indicating an approach where information is not stored into one specific location.

The approach taken is also attribute-oriented and based on the observation that information stored in hierarchical structures already has a number of descriptive attributes associated. The document `overview.pdf` stored into a tree structure carries some information already.

```
/work/projects/nexus/paper/overview.pdf
```

Each level in the structure contains information about the particular document. That it is work-related, belongs to a project, the name of the project is nexus, that it is a paper, and that the type of document is PDF. Additionally each file typically have some system specific properties as well. The Placeless Documents approach allows multiple attributes to be associated with documents in existing file systems and mechanisms for grouping documents in the system based on these attributes.

The Haystack project is also an approach that aims at providing a platform for personal information management. The focus is mainly the increasing quantities of electronic information that is dealt with on a daily basis, such as, emails, bookmarks, and electronic documents. The approach is also using an attribute-oriented scheme based on the Resource Description Language (RDF) [MM03]. Similar to the previous systems the approach allows multi organizations of information.

Lifestreams (discussed in Section 4.2.5) can also be categorized as using an attribute-oriented approach. The system allows users to search by means of keywords and substreams of matching information entities are created based on such search queries. Thus the basic mode of operation in the system is streams of documents.

In the discussion of the various attribute-oriented approaches discussed above some advantages of this approach over traditional approaches in terms of supporting personal information management specific issues have been pointed out. Although these approaches have differences the major aspects and gains of the attribute-oriented approach are common. This means that the approaches address a set of common problem typically identified in traditional systems, such as, the desktop approach, hierarchical-oriented systems, and personal information

management specific software.

The essential difference compared to traditional systems is that assigning of any number of attributes typically allows multiple ways in which the information can be retrieved and used. Thereby information that can be recalled from human memory can be used in multiple ways in order to facilitate the retrieval process in the systems. Multiple ways of organizing the information also facilitates information structures evolving over time. In traditional approaches such as hierarchical organized information it can be a comprehensive task to reorganize the information and information structures. In the attribute-oriented approach information may reside in multiple structures and thus new information structures can be created over time. Therefore the information may also be used in different contexts depending on particular task or project.

In field studies of the use of traditional file systems it has been reported [BN95] that users found that it was difficult to remember file names, even the names of files created the same day, e.g. ephemeral files. In some situations it is difficult for users to come up with a useful name for an information entity. There might not exist an obvious name for an information entity due to the nature of the particular information (see also Section 3.3 and 3.4.1). Attribute-oriented systems typically do not enforce any explicit naming of information entities. Instead different types of attributes are associated with the information. This leads to a higher degree of flexibility and no need to explicitly name an information entity when no obvious name is available.

4.2.8 Associative Approach

Hypermedia-based systems are shortly mentioned due to their widespread use today. The underlying data model of hypermedia systems expresses the inter-document relationships as the basis for the information structure. As such it is a relevant model to consider for personal information management as well. Also implementations of hypermedia-based (e.g. web-based) personal information management systems exist.

One the one side it may be fairly easy to learn and use the underlying navigation model of hypermedia-based systems, as it is fairly simple. However, user interfaces built on top of the model are not necessarily simple.

Hypermedia-based personal information management systems include web-based portal systems (e.g. Yahoo!). The portals typically allow the user to manage e-mails, address book, calendar, notes, to-dos, and a set of additional services. These may include instant messaging, bookmarks, photos, news services, weather information, and more. Typically the implementations of these systems are similar to those mentioned in Section 4.2.1 where there may be little integration between the various information types. It may also be difficult to organize related information of different types together as the different types of information are accessed and handled in a compartmentalized way.

An obvious advantage of this approach in terms of computer-based personal information management is its availability and accessibility. The information is available through any machine with a web browser and a connection to the Internet. However, from a strict user interface and interaction perspective, these approaches are constrained by the usability of the web and browsers as such, meaning sub-standard user interfaces and interaction compared to traditional graphical user interfaces.

The main point in terms of personal information management is the presentation of interlinked information. Structure and organization of the information is by a graph of interrelated information. But most hypermedia systems still only have one canonical view of the information entities [Nie95]. As have been discussed multiple views of information are relevant for personal information. Also different information entities may be related and linked using the associative approach. Only some initial attempts have been made to develop personal information management systems based on such an associative approach.

4.2.9 Personal Digital Assistants

The desktop and web paradigms represent the domination application user interfaces of today for the stationary situation of use, such as, a desktop (or laptop) PC with a user sitting behind a display, using the keyboard and mouse as the input. However, in the last decade Personal Digital Assistants (PDA's) have become popular and widespread also. A direct comparison is not appropriate, since PDA's are handheld devices that often fit in a shirt pocket and is meant for a mobile situation of use.

Nevertheless PDA's are mentioned, as they are typically built around the notion of personal information management. An example is the Palm-based user interface, which is tailor-made for personal information management [Ber00] compared to modern desktop implementations. PDA's typically support tasks such as keeping calendar information, address information, reminders, notes and memos. Some PDA's support additional personal information management related information, such as, e-mail and instant messages, expenses, et cetera.

These type of systems are typically similar to those mentioned in Section 4.2.1, however mobile information appliances, such as, mobile phones and PDA's have limited functionality compared to desktop systems, due to lower computing power, and limited input and output capabilities. Input is provided trough a pen or a small-sized keyboard, and output through a display with smaller size and resolution. The situation of use is different, as the user typically holding the device with one hand and operating it with the other (pen or keyboard). Generally these systems are targeted for a mobile situation of use, which is of course highly relevant aspect of personal information management too.

4.3 Discussion

In the previous sections a wide range of different approaches to the support of personal information management by means of an interactive systems have been discussed. The approaches are somewhat different and each emphasize and address its own set of personal information management problems.

The following sections point out some of the essential issues in personal information management and discuss to which extent the different approaches address these issues.

4.3.1 Structure Part of the Interface or Part of the Information

The organization of information is obviously one of the most essential issues in personal information management, that is, how the user may structure his personal information by means of a personal information management system. As discussed in Section 2.7 different levels of organizational structure is essential

as the different classes of personal information typically dealt with is structured in different ways. For instance the level of structure applied to information depends of the use of the information.

Both unstructured, semi-structured, and formally structured information can be found in the way humans deal with and organize their personal information. Typical systems only include an information organization scheme that is based on the hierarchical structured information. An example is email software applications that allow the user to perform an organization of emails into a hierarchical structure only. This approach is suitable for formally structured information, but is less suitable for less structured information. Other approaches support less structured information, for instance the pile metaphor from apple [MSW92] and some attribute-oriented approaches as well (as mentioned in Section 4.2.7).

One issue is the problem of varying levels of structure. As mentioned the approaches might allow more or less flexible ways of organizing information at various levels of structure. Some systems determine the structure of information as it is built into the system (the user interface). The email software mentioned above is an example of this approach. Other systems enable the user to decide the actual structure of the information rather than the system.

In traditional record-oriented databases, such as relational databases, systems developers are the ones that create and change the structure of the database (rather than end-users). Once defined a record-oriented database has a static structure into which the user has to organize his information. A developer rather than an end-user of the system does changes of the information structure as such. As has been discussed previously personal information management is characterized by information that is sometimes poorly structured and in an ad hoc manner [Mal83] [KKB+90]. As such the context and structure of personal information is likely to evolve and therefore change over time. Many of the traditional systems mentioned in Section 4.2 are inflexible in this way as the end-users are restricted by the structure available in the system. Some systems require that a systems developer is capable of adapting the database to current needs. Other systems allow end-user to adapt the structure of the information over time as the context of the information and the structure need to change [KKB+90].

The general issue is whether the *information structure is part of the user interface* of a particular system or the *information structure is a part of the*

information (the content).

Another example of the first type is hierarchical-based organization, which is somewhat inflexible as the structure is enforced by the user interface of systems based on this approach. For instance file systems where the structure is an inherent part of the user interface rather than the information itself. The user typically has no other option than to organize the information (e.g. files) into the hierarchical structure provided by the user interface. Even if the information to be organized is not of a hierarchical nature it still has to fit into that particular information structure. Thus the structure is part of the user interface and less flexibility is left to the user in terms of information organization.

In the latter approach the information structure is part of the information itself. This means that the structure of the information is determined by the user by whatever means is available in a particular system. This means a higher level of flexibility is provided for the user to organize the information. In this case the structure becomes part of the information itself rather than being predetermined by a particular system.

Table 4.1 provides an overview of the approaches previously discussed and indications of the information organization approach used in the particular systems.

4.3.2 Ephemeral, Working, and Archived Information

Another essential personal information management issue different levels of interaction with personal information. This issue is related to the issue of various levels of structuring of the information discussed in the previous section.

As discussed in Section 2.7 three classes of information or levels of interaction has been identified in studies of personal information management. This has been identified in both physical office environments and in the use of interactive systems for personal information management. However, the level of support for these different types of information varies widely in existing personal information management approaches.

The support of *archived information* has received a lot of attention in both research and existing systems, such as information retrieval and database systems

	Structure part of content	Structure part of the interface
Lifestreams	Entities have associated attributes	Time-based structure
Haystack	Collections formed by properties associated with objects	
Placeless Documents	Collections formed by attributes associated with documents	
Lotus Agenda	Items have category attributes and user-defined views	
PAD++	Organization is decided by the user (in the spatial layout)	
Desktop		Hierarchical file structure
Palm OS		Predefined calendar and lists

Table 4.1: Information structure in five different approaches that support personal information management
1) Information structure is part of content, or 2) information structure is part of the user interface of the system

for long-term storage (archiving) of information. All approaches discussed above provide support for archiving information and aim to facilitate the process of archiving information for later retrieval. The mechanisms to achieve this aim are different in terms of hierarchical-based system, temporal ordered information, or spatial arranged information. For instance desktop-based systems with the filing cabinet metaphor also primarily provide support for archived information, and less support of the other levels of interaction.

Working information is characterized as active information that is being used as part of ongoing activities, and therefore the information must be easily accessible in the immediate information environment. This level of interaction has some level of support in personal information management systems. For instance

the desktop-approach provides multiple open windows that provide multiple simultaneous views of active information. This enables multiple concurrent access points to working information, and an analogy to a physical desk with multiple personal documents used as part of an ongoing activity. Another way the support is achieved is placing active information on the desktop. The desktop environment allows spatial layout of the active information thereby providing access through the immediately available environment. In the study by Barreau and Nardi [BN95] subjects reported that this class of information was supported sufficiently in the desktop-based system used in the studies.

Supporting *ephemeral information* has received little attention in human computer interaction research and in the design of personal information management systems according to Nardi and Barreau [BN95]. Among the different levels of interaction the primary lack of support seems to be in the area of ephemeral information. Especially handling large quantities of ephemeral information is still an issue that needs further research. Ephemeral information is characterized as being unstructured, or semi-structured and used in an ad hoc manner. Current systems (e.g. file systems) demand that information entities are provided with a name and stored into a hierarchical structure. This is not suitable for ephemeral information due to its characteristics. It must be easy and fast to get a piece of ephemeral information into a system, if it needs to be handled in the system. If a system requires multiple steps to create a document, provide a name, before the information can get into the system it would seem too troublesome to handle this type of information.

The support of the different levels of interaction varies among the different approaches different sets of personal information management problems are addressed by the approaches. However, only limited studies have been carried out to in order to get an understanding of how the various levels of interaction are supported in the different approaches. Barreau and Nardi [BN95] have done some initial observations in studies of file organization using desktop-based systems. The level of support for the different levels of interaction is quite different in the general approaches. This is due to the emphasis on different aspects of personal information management. Especially archived information has received attention. However, further research is needed in order to establish an understanding of actual level of support in the different approaches to personal information management.

4.3.3 Effect of Approach on User Strategies

The approaches to the support of personal information management considered in this chapter are different and also address different sets of problems. The approaches support different aspects of personal information management, which also means that at some point there could be implications of the actual approach or system on the information organization and management strategies chosen by the users. The question is to which extend the system as such has implications on the personal information management strategies.

Fertig et al. [FFG96a] have argued that the tools available[1] could strongly influence and constrain the filing and organization strategies chosen by the subjects. They argue that some of the findings of Barreau and Nardi [BN95] are artifacts of the desktop and file metaphor (hierarchical filing scheme). However, similar findings to those of Barreau and Nardi have been reported from other studies of information organization [Mal83] [Col82] [Lan88].

Nevertheless, the argument is relevant, as it seems to be a reasonable guess that there would be influence from the supporting system on strategies chosen, especially in terms of the constraints of the system. Additionally the lack of some features in an interactive system might influence the users to choose non-optimal filing strategies. For instance the reluctance to use search features might be caused by their inefficiencies [BN95]. Users might develop different filing preferences if the system implemented efficient indexing schemes, thereby drastically reducing the search response time. It is often faster to locate a file (document) using modern web search engines than finding a file using the standard search mechanism on modern desktop systems in terms of system response time.

Fertig et al. [FFG96a] also question the preference for and value of location-based search in terms of the periods of time it is used. The way the information is used might change over time. Therefore Fertig et al. suggests using the temporal dimension as the key to information archival in their Lifestreams model. However, as have been discussed information retrieval is not one-dimensional, but based on associations with information of similar type, topic, time or place [FFG96a]. It has been pointed out in several studies, that location is important for the reminding function, however, Fertig et al. suggests that the

[1] Fertig et al. discuss filing using desktop systems and hierarchical filing systems

observations in the study by Barreau and Nardi might be a coping strategy since the systems studied did not offer any better alternatives. It is considered an ad hoc user convention, which has a number of potential problems. For instance there is no guarantee that the user will actually be reminded and that it only scales to a limited number of reminders, due to limited screen space on typical systems. Nevertheless, there are indications from multiple studies that users do in fact use location as a reminding tool, in one way or another. This has been identified in both physical environments [Col82] [Mal83] and interactive environments [BN95].

A conclusion from the study by Barreau and Nardi [BN95] was that subjects were reluctant to archive information. It is pointed out by Fertig et al. [FFG96a] that this might be a consequence of the systems used by the subjects in the study. The systems force users to come up with a filing scheme, and organize information (files) into a hierarchical structure. As discussed previously, this is a time consuming and difficult task. The effort to create the scheme may exceed the gain.

Additionally the context in which the information is used changes over time. This might explain the reluctance of the subjects to archive their information. Barreau and Nardi [BN95] reported how subjects explain how they have developed elaborate filing schemes, but abandon using them since it required too much work to maintain and use them. It seems that an archiving scheme may only be suitable for a limited collection of personal information; it does not scale beyond a certain quantity of information or beyond a certain time span. Either the filing scheme need to evolve as the quantity of information increases, or this could happen over time as personal information needs changes.

4.3.4 Summary of Approaches

The number of issues involved in personal information management makes it quite complex. The different approaches have different levels of support for the various issues involved. The previous chapters have provided an attempt to summarize the essential issues involved in personal information management and the human cognitive issues involved. In this chapter the issues have been discussed in terms of different system approaches to personal information management. Each

approach typically emphasizes its own set of features in terms of supporting and facilitating personal information management.

In the desktop environment each application allows the user to create and edit a file, but with no notion of the context in which the file is used. The files are organization into a hierarchical file system, and the user must create a classification scheme, and naming of files. Additionally, the user can use the desktop as a spatial placeholder for icons that are references to files in the file system. Thus the model can be regarded as being a hybrid hierarchical, spatial, and semantic (naming/identifying) information structure. Reminding is typically supported by means of leaving information in specific locations, such as the desktop or in the email inbox [ABC98].

The Lifestreams approach addresses the problems of naming and organizing, as all documents are automatically temporally organized with the benefit of no need to explicit name or organize documents. The emphasis is on temporally structured information, and as a consequence relies on the ability to retrieve information by establishing an earlier context by means of recalling the temporal context of the information.

Spatial approaches, such as zoomable interfaces utilize the human spatial memory capabilities and a spatial oriented way of organizing information. These approaches also address the problems of explicit naming of information entities. However, a potential problem is that information may only reside in one physical location, when it may belong to multiple contexts and groups of information. Simply copying the information to multiple locations introduces other problems, such as keeping all copies up to date.

Specialized tailor-made user interface for personal information management for handheld devices such as Palm devices. The information structure is based on one-dimensional (one level) categories, as each information entity can be associated with exactly one category.

The addition of piling capabilities into existing systems, adds the capabilities of semi-structured information which otherwise has little support in hierarchical information archiving systems. This approach also avoids explicit naming of information entities.

Attribute-oriented approaches such as the ones found in Lotus Agenda, Placeless documents, and Haystack, emphasize the different levels of structure of personal information. It allows unstructured, semi-structured, and formally structured information. The information structure is based on a one-to-many relationship between information entities and associated attributes. Such approaches allow the structure to change over time, and multiple organizations of the same information.

Table 4.3, 4.2 and 4.4 provide an overview of issues involved and indicate the support provided in the different general approaches to the support of personal information management that has been identified in the research of the area.

4.4 Summary

Given the human capabilities, such as memory capabilities, the challenge is to construct systems so that they support the specifics of personal information management. Based on what can be recalled from memory intent is expressed. This can be to organize, retrieve or otherwise use information kept in an external representation. The problem is to express this intent based on recall of personal information, and easily map it to actions that are useful in order to organize, retrieve or use information in the external representation.

As discussed in the previous chapter the humans are capable of recalling information along many different dimensions. However, as seen in this chapter many current approaches to personal information management only allow a fixed and limited way of accessing and organizing information. This forces a transformation of what the human is capable of recalling about the information into a specific format or request in the system that holds the external representation of the personal information.

Often humans are capable of expressing a lot more about their (personal) information than can be mapped directly onto an external representation of that information in a system that aims to support and facilitate personal information management. A challenge is this gap between expressed intent by the human and which cues can be used to organizing, retrieve and use information in an external representation of that information.

In addition the issues involved in personal information management are quite complex, as it involves different classes of information, with different levels of interaction, organized in different ways, and where information use and information structures are likely to change over time. This means that personal information management systems need to support a diversity of activities and diversity in ways the information is handled. The different approaches to the support of personal information management discussed each emphasize different aspects of personal information management and therefore also provide different levels of support of the issues involved.

Some problems identified in current systems can be ascribed to the fact that the information environment of present day is somewhat different from the information environment these systems and approaches were designed to deal with. The discussion of personal information management support in Section 4.2.1 briefly mentioned the issue that personal information environments are changing as more electronic information is used as part of daily activities. This includes information types such as e-mails, bookmarks, and electronic documents.

Originally the desktop metaphor of the Xerox Star [JRV+89] and Apple Macintosh was meant to do office automation and provide an easy understandable and quickly learned user interface. It is based on a document-oriented information approach mapping existing offices practices to the desktop system supporting tasks and applications such as word processing. The hardware of that time was only capable of storing a smaller quantity of documents on local hard drives. Few applications available each with few commands made the available information and functionality manageable. Today systems based on the desktop metaphor are used to support personal information management [BN95] although that was not the primary aim. The user interface therefore supports handling documents that can be printed, rather than focus on the specific issues involved in personal information management. Such systems designed as filing cabinets and for printed documents does not address issues such as piling information, or ephemeral notes, or tagging and dog-ear information.

Some of the changes that have occurred include:

- The number of different applications has increased
- The functionality in terms of the number of commands in each new application version has increased
- The types of different information has increased (e.g. multimedia content)
- The quantity of information handled has increased
- Connection to networked resources has contributed drastically to the increased quantity of information available
- Communication through interactive systems (e.g. email and instant messaging)
- The number of different information appliances used to support information management has increased yielding a more complex environment of devices and information repositories

These changes have lead to an increasing volume and variety of the ephemeral and working information that the user must potentially manage.

The approaches discussed in this chapter provide different levels of support of the aspects and activities involved in personal information management. It is a considerable challenge for systems to support this diversity of aspects and activities involved in personal information management. In several studies within the field of personal information management there is an agreement on a call for better tools for the support of personal information management. Especially for the support of the increasing volumes and variety of ephemeral and working information have been pointed out [BN95] [FFG96a] as this has received less research attention. Moreover as the complexity of the information environment, available information, and types of information increases the need for a fundamental understanding of the processes involved in personal information management increases. This understanding could potentially provide the basis needed in order to provide such tools.

Approach	Metaphor	Informa–tion entity	Organization of informa–tion entities	Naming of information entities	Information access and entry
Life–streams	Diary or pile	Docu–ments	Temporal streams	No – attributes can be associated	External appli–cations
Haystack	None	Objects	Views – properties associated	No – properties associated	Integrated in views
Placeless Docu–ments	None	Docu–ments	Collections attributes associated	No – attributes associated	External appli–cations
Lotus Agenda	None	Items	List in user–defined views	No – attributes can be associated	Integrated in views
PAD++	"Spatial"	Docu–ments	Spatial	No	External appli–cations
Desktop	Desktop	Files Docu–ments	Hierarchical file system	Files and folders have names	External appli–cations
Palm OS	None	Entries (text)	Application specific (lists)	No – one category can be associated	Through "appli–cations"

Table 4.2: An overview of properties in different interactive systems that support personal information management
1) The metaphor the system (user interface) is based on, 2) The information unit of work, 3) The organizational principle for information entities, 4) Naming of the information entities, if any, and 5) How information access and entry is provided

Approach	Spatial information organization	Temporal information organization	Information classification	Finding information by content
Lifestreams	By the temporal dimension	Files ordered by time in "streams"	Associated attributes	Search feature over entire set of documents
Haystack	Spatial layout of information views	Views ordered by time	Associated properties	Search facilities
Placeless Documents	Yes	No	Associated attributes	Spatial and in collections
Lotus Agenda	No	Entries ordered by time (view)	Items can have multiple categories(1)	Search feature across entire set of entities
PAD++	Yes	No – only if temporally ordered by the user	Levels in spatial organization	Search spatial layout
Desktop	Icons on the desktop	Sort folders by time	Files in folders hierarchy	Search folders, file system, or drives
Palm OS	No	Implicit in calendar only	Entries are assigned only one category(2)	Search feature across entire set of entities

Table 4.3: Information entity properties in five different approaches to the support of personal information management

(1) The categories in Agenda are organized in a hierarchy. An item associated with a category is also automatically associated with the parent category.

(2) The categories in the Palm OS are application specific. This means that the calendar has one set of categories and the address book has a different set. Thus the user has to maintain multiple parallel category sets.

Project	Spa–tial	Tem–poral	Attri–butes	Hier–archy	Graph Graph	Unstruc–tured
Lifestreams	(+)	+	+	–	–	(+)
Haystack	+	(+)	+	?	?	?
Placeless documents	+	–	+	(+)	(+)	(+)
Lotus Agenda	–	(+)	+	–	–	(+)
PAD++	+	–	–	(+)	–	+
"Pile" metaphor	(+)	(+)	–	–	–	+
Desktop systems	+	–	–	+	(+)	(+)
File System	–	(+)	–	+	(+)	–
Palm OS	–	–	(+)	–	–	–

Table 4.4: Types of information organization supported by different approaches
+ supported, (+) partly supported, – unsupported,
and ? indicates that is it unclear to what extent it is supported

Chapter 5

Unification

This chapter extends the discussion of support of personal information management to include user interface and interaction specific issues. This discussion includes user interface and interaction oriented problems that have been identified in studies of personal information management and that have implications for the support of personal information management.

The previous chapter discussed how the different existing personal information management systems provide different levels of support for the issues involved in personal information management as such. From a user interface and interaction perspective it is discussed how it has been observed that humans come up with different coping strategies in order to carry out their activities and manage their personal information. This includes interesting observations of unintended use of the systems in order to obtain the needed support of personal information management related activities.

Prior research of personal information management includes studies of information organization practices in different domains and research of different existing and novel systems that aim to address different issues in personal information management. Often these studies focus on some kind of improvement for a single specific information type. For instance there have been studies of enhancements for email management [BDHS03] or bookmark management [RCL+98].

Although personal information involves many different types of information

and activities few studies have taken a general integrated approach to personal information management. Although some types of information might be handled in an isolated way different types of personal information is often handled together by humans. However, as discussed in the previous chapter not all personal information management systems enable such combination, aggregation, or interconnection of disparate but related information entities. Such personal information includes mail/messaging (asynchronous and synchronous communication), to-do items (priority of tasks), calendar information (temporal-based information), addresses, documents, et cetera. Rather some systems have a compartmentalized way of supporting these different types of information. Each of these has different modes of operation, which makes it difficult to combine and organize related information. Moreover it means that users have to learn and use multiple modes of operations depending on the information at hand, even though the way the information is management is identical (e.g. in a hierarchical structure).

A simple example of inconsistent access to related or similar information in system is in the way current systems enable access to information that is kept in a local system versus the way in which information is accessed when it is located remotely on a server. For instance current systems provide a desktop and file system for accessing local information, while for instance the webbrowser is used to access remote information. Therefore the user experience is two different models to support information from different locations, even though essentially the same information is retrieved. As such two different metaphors are living side by side to do essentially the same [Der01]. For example Raskin has suggested a unified user interface to local as well as remote information (available through a network) [Ras00].

Finally, as mentioned in the previous chapter, current systems are used for a considerably larger variation of functionality, applications, services and information. The kind of information that is handled with computers are no longer limited to text documents, but now include spreadsheets, images, e-mails, notes, calendar entries, bookmarks, sound, video, and other application specific information. In addition network delivered content drastically increases the quantity of information that needs to be managed by humans. However, the same underlying model is still used as the basis for information management – even if the numbers mentioned above have increased drastically.

5.1 Usage of Existing Systems

Studies of personal information management have identified problems in current systems and that existing personal information management systems do not address the complexity, variety of information, and activities in personal environments. Additionally some interesting examples of unintended use of systems and strategies used to cope with limited support of individual style or practices have been observed.

5.1.1 Use of Email Systems

There have been some interesting findings in studies of the use of email. One such finding is that some users prefer to use their email system as a central repository and manage a large subset of their personal information through the email system. In these cases the email system has become the main tool for personal information management. This means that they use the application for more than just sending, receiving, and organizing e-mails. For instance they also use it for keeping reminders, task management [Mac88], document management, and archiving [WS96]. Some also use the email system to keep bookmarks to webpages and other material [JBD01].

These findings have lead to research that have tried different improvements of email systems in order to enhance the support of these various types of information and activities. A recent example is the Taskmaster approach [BDHS03]. The combination of the different types of information and activities in the same tool is an indication that messaging, task management, to-do items, and other information is used together as part of the daily activities, and therefore naturally put together in the same context. This indicates a need to keep related information "in one place" and having "a central point of access", that is, the possibility of interrelating and organization of personal information across information types in the same information structures. Therefore it also seems insufficient to improve access and support for only one content type when personal information management involves a variety of different information types that are used together and across different daily activities.

5.1.2 Keeping Bookmarks

Another interesting finding is from observations of keeping bookmarks to webpages [ABC98]. Typical webbrowsers have a built in feature that allow users to keep bookmarks to a webpage for later retrieval. Typically the URL, the title of the webpage, and the date of retrieval are stored with the bookmark, and this information item is typically kept in a hierarchical information structure in the webbrowser. This feature has been specifically designed to support users in keep track of webpages. Due to the hierarchical structure a problem is that as the number of bookmarks increases beyond a certain level it becomes difficult to manage the bookmarks and less efficient to find bookmarks among the others. Empirical studies have shown that users do not generally use the bookmark features in webbrowsers at all [JBD01]. The primary reason being that the bookmark system does not fit into the tasks that are being carried out. This means that the bookmark feature does not store bookmarks in the context in which they are used. As an example a bookmark does not serve as a reminder if put into the bookmark system as the users will have to explicitly visit the bookmark system to obtain such reminders.

Participants in studies of the use of bookmarks [JBD01] used a wide variety of different strategies in order to keep such information, and make use of it with other information and as part of other activities. For instance ad hoc methods were used to cope with the lack of the reminding feature. I situations where reminding was needed participants would instead e-mail a bookmark to themselves or leave an icon on the desktop to serve as a reminder. Many additional strategies have been identified in empirical studies of bookmark management. Examples of strategies include:

- Storing bookmarks in documents
- Printing pages
- E-mailing bookmarks to them selves
- Storing bookmarks as shortcuts in file structure
- Storing bookmarks as shortcuts on desktop

- Storing bookmark in personal website
- Using bookmark feature in web browser
- Typing in the reference directly

The examples of non-use of a feature that have been designed specifically to serve the purpose of keeping track of webpages indicates that it clearly does not fulfill this needs of the participants. The information is used in a context and the bookmarks systems in the studies did not allow the participants to manage bookmarks in the context in which they were used. This explains the variety of coping strategies observed that allowed them to manage and use bookmarks in the context of use. In the example where participants send bookmarks to themselves in an email enable them to have contextual information with the bookmark, and then the bookmark (in the shape of an email) serves a reminding function.

Other studies indicate that the same set of problems exists whether users organize documents, emails or bookmarks into a hierarchical structure [WS96] [JBD01].

5.1.3 Multiple Parallel Structures

A somewhat related finding is the result of studies carried out by Boardman [Boa01]. A number of subjects were interviewed about their information filing practices, and in particular their development of information hierarchies. It was observed that some participants used similar categorization schemes and had similar information structures in multiple distinct applications. As an example they would have same hierarchical information structure in the file system and e-mail folders. This means that the different information hierarchies in the personal information space had similar structures in terms of category overlap, for instance in management of email and documents. The categories were overlapping as category labels were based on projects and roles. It suggests that one scheme may be used across the entire set of information types.

In order to keep such multiple structures the problem is that users must perform many equivalent tasks in different applications multiple times in order

to manually keep the multiple structures consistent and up-to-date across the different applications. This is needed in order to organize the different types of information used into the same (parallel) structures. Besides managing parallel structures and keeping them up-to-date, each application may have inconsistent commands that operate on the information hierarchy. The suggestion in the study is that this manual maintenance of multiple parallel information structure should somehow be automated in the system.

Boardman [Boa01] developed a solution where the information structures are automatically updated by the system across different applications. Meaning the system makes sure that the system uses the same information structure in different applications. For instance the hierarchical information structure is the same for organizing emails, bookmarks, and files in the file system. Although this approach address the problem of manually updating and keeping a similar information structure across application domains, it still does not address the issue that information is kept in different compartmentalized areas in the system. As such the system does not allow the user to store information of different types into the same information structure, as emails are kept in one part of the system and bookmarks in another. The approach only allows the hierarchical structure as such to be kept the same across these different parts of the system.

5.2 Modality

Existing systems often have multiple modalities and modes of operations, which lead to confusion and that users have to learn multiple sets of commands to manage information across different compartments of the system. Even though the underlying type of information management is the same across different applications in desktop-based system still different commands sets have to be used.

5.2.1 Different Contexts

As such the problems of performing personal information management and retrieval show up in a number of different contexts. To a large extent the problems can be considered to be equivalent, as the same problems of information man-

agement and retrieval are found in different contexts in the tasks performed by users of current systems. Even though the information entity types are different the same set of problem apply in information management in different situations. Some typical examples from the desktop environment include:

- Managing files (documents) in a hierarchical file system
- Managing e-mails in an e-mail program (with hierarchical folder structure)
- Managing bookmarks to webpages in webbrowsers (in hierarchical structure)

These systems are common in that they typically allow management of application specific information only and therefore not managing and relating different types of information. This is in spite of the fact that each system uses the same underlying approach for information organization – a hierarchical information structure. The previous chapters have already discussed the potential problems such as the cognitive burden put on the user to recollect where the information is located in the structure. The additional problem is the compartmentalized way of supporting these different types of information.

Even though these information entities are of different type the problem of categorization or placement in a hierarchical structure are the same [WS96] [JBD01]. It has to be done in inconsistent ways even though the user performs similar tasks. This is caused by the hierarchies being managed in multiple distinct applications. Each application has its own way of implementing the hierarchy and commands to use it. This means inconsistency in the user interface environment, thus making it difficult to learn and remember to use. Non-unified commands and inconsistency leads to additional cognitive burden for the user, due to the extra effort learning a different way of navigation and using parts of system based on the information types.

5.2.2 Overlapping Functionality

Modern implementations of the WIMP model typically make available more than a handful of overlapping functionality in the applications domains. For instance

editing text is typically available in different parts of the system. E.g. there are editors in word processor, e-mail application, spreadsheet, presentation program, notes, et cetera. This requires the user to edit and manage personal text information in different inconsistent ways depending on the application used. Therefore different modes of operation exist in the user interface, that is, different ways to do conceptually the same thing (in this case editing text). An advantage is that there may be needs for tailoring functionality and available commands for specific situation of use. However, the disadvantage is that the user has to learn and remember multiple ways to basically do the same thing across different applications.

There is only limited unification of commands in desktop systems. Designers try to achieve it by having design guidelines. If followed the purpose is to lead to greater level of consistency in the user experience – even when using a set of different applications from different application vendors. Assigning the same short cuts – assigning menu items to the same location, and generally aim for similar menu structures. Inconsistencies are inevitable in practice with different vendors of software applications.

Figure 5.1 shows a Venn diagram illustrating the overlap of commands in typical desktop applications, such as, (a) word-processing, (b) spreadsheet, and (c) graphics presentation. Specifically the commands dealing with the management of information entities are common among the different applications. The information dealt with by the application is organized in a common information structure. As mentioned in the previous section there are different compartments handling the different information types using different sets of commands.

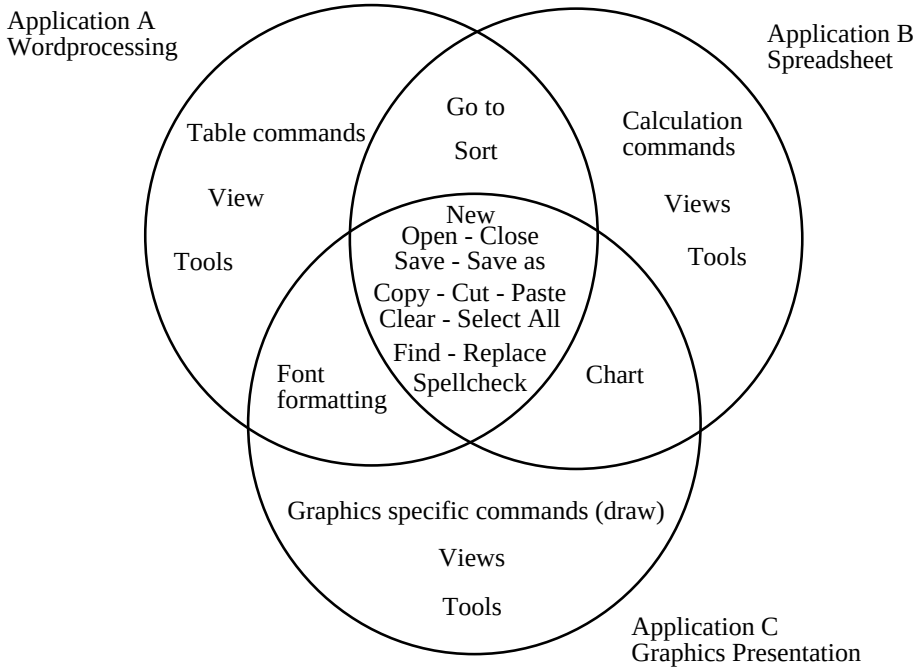

Figure 5.1: Illustration of overlapping functionality in typical desktop applications

5.2.3 Modes and Consistency

The operations that apply on a hierarchical information structure are essentially the same across different information types. Such common operations include:

- Create or delete an information item
- Create or delete a node
- Move information item or node (subtree)
- Expand or collapse part of the structure

Different systems typically use different commands with different names in order to perform these fundamental operations on the hierarchy. This means that the user has to learn and use different sets of commands that fundamentally do the same. For instance one set of commands in the email system, another for managing documents, and yet another to manage bookmarks.

Creating, deleting, moving, rearranging information entities are conceptually the same operations, independent of the applications used. It is the same set of basic commands available in the different applications or modes. Location of commands in menu, and keyboard shortcuts for those commands is what can make the user interface consistent. Such consistency is not a guarantee. It is only the case if designers of the systems have followed a set of user interface guidelines.

Inconsistencies may lead to confusion and mistakes. As the user interface may also be in different modes the user need to be aware of which mode is the current mode of operation. Otherwise this may cause mistakes and unintentional changes of information. The different modes in the user interface may cause difficulties for the user during use. Also the user will need additional effort to learn the different modes of operation as multiple sets of commands are used.

One potential advantage of the compartmentalized approach where functionality are separated is that it makes it easy to visually distinguish different "applications" and information types, since the presentation of information may be different due to the different modes or applications. There is a clear mapping of information types to applications, which means that it is clear to the user that emails are found in one compartment and documents in another compartment of the system.

As mentioned above the disadvantages of the approach is that it makes it difficult to learn and use several different modes of operation. Users need to be aware of the mode, which means less attention to the actual task being carried out. Additionally the separation and multiple modes of operations may lead to confusion and possible mistakes.

Related information cannot be combined in the context of use. From a personal information management perspective the compartmentalized approach might be hindering natural ways of combining and aggregating related information entities. Also in different contexts of use information may need different

levels of formality in organization. Thus information entities may naturally appear as part of different organizations depending on the situation of use. For instance an email may serve as a reminder, or may be part of active ongoing tasks, or may be archived for later retrieval.

5.2.4 Granularity of Information Units

In Section 2.8.1 it was mentioned that the unit of work identified in the study by Malone of paper-based personal information management in an office environment [Mal83] were *files* and *piles*. Both help collecting groups of elements into larger units. Files are explicitly titled and arranged by some criteria. Piles are not arranged in a systematic way, why their spatial location is important for finding. This approach has been partly mapped to desktop-based systems, which have files and folders.

These abstractions represent levels of information granularity that are used in the systems. Thus the information units are the unit of work in the system. The level of granularity of information units has implications for the flexibility of management and use of the information in a system.

Desktop-based systems have course-grained information granularity in the shape of files that represent application specific information units, e.g. documents. Thus each information unit contains a specific type of information that is represented in a specific way. It may contain a complex information structure that can only be managed and organized in an application specific way where the application has a predefined structure for the information. This means that the structure is defined by the system in advance rather than being defined or organized by the user of the information (see Section 4.3.1 information part of the structure or part of the interface).

This course-grained processing of information (files) in desktop-based systems have implications for the way in which information is organized in such systems. The underlying model is information processing of a file that serves as input, an application doing processing of that file, which finally generate output of a file. This means information is retrieved from a file structure, processed, and then stored into the structure. Thus the basic metaphor, that is, the simulation of processing of documents in an office. This approach has limited support of per-

sonal information management in terms of the variety of ephemeral and working information where unstructured and semi-structured organization is preferred. The explicit application commands for loading and saving of files from and into the file system may be inconvenient for this type of information. For instance Raskin [Ras00] argues that no user interfaces should ask the user to explicitly perform a "save" command. The system should automatically persist the users work so that he does not have to think about it. Modern designs as for instance in the Palm Operating System [Ber00] has altogether abandoned explicit user actions for saving information, thus proving by example that it is not needed in modern user interface designs.

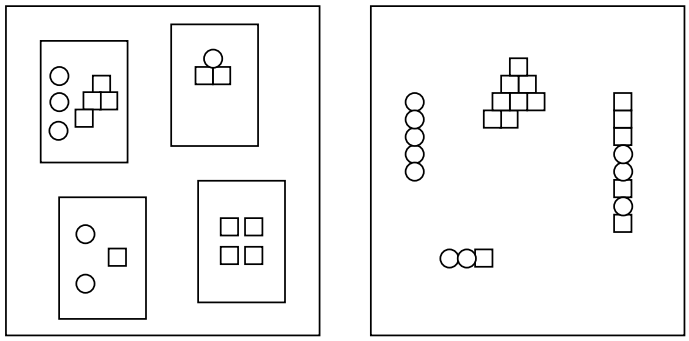

Figure 5.2: Illustration of systems with coarse and fine information granularity In the first information elements are parts of larger structures (e.g. specific system-defined structures). The other system has fine information granularity where different fine granularity information entities can be organized in arbitrary information structures

Additionally interaction based on such course-grained granularity of information units may lead to information and functionality being compartmentalized as described in the previous subsection. The problems include different information units in the system having different behavior and need to be used in different ways due to different and perhaps inconsistent modes of operation. In terms of personal information this problem is illustrated in specific personal information management systems where different information types are organized and used

in separate compartments of the system. This imposes a set of limitations to the way in which information can be organized and used. Additionally there are limitations to the flexibility in terms of information structures. Figure 5.2 illustrates the differences in the approaches. It also illustrates the difference between information structure being part of the user interface versus being part of the information, as discussed in Section 4.3.1.

5.3 Specific versus Generic Tools

The issue of granularity of information is related to the issue whether generic or specific tools provide the best support for user activities, such as personal information management related activities.

5.3.1 Flexibility in Information Entry and Use

Many personal information management software applications typically include personal information management functionality such as, calendar, contacts, to-do, notes, messages, et cetera. These different functional aspects are sometimes integrated in terms of linking related information. However, information and user interface is often compartmentalized, as each information type is handled in a different way as discussed above. Also when creating an information item, the user has to decide in advance of "which type" the information is. An information item may start out as a brief note, but may turn into a to-do item, an e-mail, and so on. Therefore it may also be used in different contexts, and the information item may be part of multiple information structures.

In the physical world you could take a blank piece of paper, and you do not have decide a priori what you are going to write on in or use it for. It may be for writing a brief note for a shopping list or to do list. It is flexible as it can contain anything and combinations of personal information that is needed to support a user task. Additionally it is easy to write annotations anywhere on the papers, for instance in a calendar system. Entering information into an interactive system, means that perhaps fields of data have been decided in a fixed format, and the type of data in the fields are decided too. This provides less flexibility compared to a paper-based system.

Examples include:

- An e-mail received could contain comments to a project proposal, why it is not only an e-mail, but also has relevance as notes to a project.
- An e-mail received may include the scheduling of a meeting. Thus the information may be turned into a calendar item including the meeting agenda, and the persons participating in the meeting.
- An e-mail received may include information about a work task to be carried out. Thus the information may be turned into a reminder of a task to do.

In all the examples above the information originate from an e-mail received (communication), and exemplify how such information may turn into other types of personal information, which are related to different activities. The observations from paper-based support of personal information management and these examples indicate that different functionality and ways of organizing the information is important in order to provide the flexibility needed in order to support the variety of information and processes involved in personal information management.

5.3.2 Example: Unix Pipelining

An example of a system that provides a high level of flexibility in terms of combining functionality and allowing the user to manipulate information in flexible ways is command shells of UNIX-based operating systems. Specifically the scripting and so-called "pipelining" capabilities provide this flexibility. Pipelining enables the user to "pipe" information streams between different programs in order to obtain a goal. Input from a file can be piped as input to a program. The output from that program can be piped as input to yet another program or written to a file. A simple illustrative example is shown below:

```
cat input | sort | uniq > output
```

It uses the file `input` as the input, piping it to the `sort` command that alphabetically sorts all lines from the input. Then the sorted data is sent to

the `uniq` command that removes duplicate lines and finally the result of that process is written to a file called `output`. In this trivial example three different tools (commands) are combined to achieve the goal. Each tool can be used in combination with any other tools available in the system. If the user needs to perform this specific combination of commands often they can be written into a script. That makes a user-defined command available in the system that can be used like any other command available. If repetitive tasks are needed they can also be written into a script. However, as mentioned above this kind of user interface requires that the user is willing to invest the considerable time needed to learn the commands available in order for the user interface to become efficient to the user and for the user to achieve this level of flexibility in the user interface. Each command typically has many parameters, which are not easy to remember, if not used often. UNIX operating systems typically contain a fairly large set of such low granularity tools that can do specific transformations of the input. A reason why this model is very useful is that the data type is typically text only. A sequence of text lines is the standard data format used. Even though the model as such is simple it is powerful, but not targeted for the average user.

Even though the UNIX command and pipeline paradigm is very effective is has received critique because of the cryptic command names, and syntax, which is very hard to learn and use [Nor81]. However, for the narrow domain of users, such as programmers and developers, it is widely popular for the efficiency. It is included as an illustrative example of the differences between a specific and a generic approach to information handling and use.

Another advantage of this approach is that it enables easy repeatable tasks, whereas for instance the desktop approach is insufficient in handling repetitive tasks. Approaches that address this issue and add the capability to desktop systems has been researched [Cyp91] and [Lie98]. In their approaches agents monitor the system and tries to automatically detect interaction patterns.

5.3.3 Studies of User Preferences

Nardi and Johnson [NJ94] [JN96] have carried out ethnographic studies of user preference for task-specific versus generic applications. The research was based on interviews with people who used different software tools to prepare presen-

tation slides. It was studies how well the different tools supported the various aspects of the task of creating presentation slides. This issue of task-specific or generic applications for the support of user tasks is relevant in context of personal information management as well. As mentioned above some personal information management systems are specific tools that support a specific subset of personal information (calendar, to-do items, addresses). Others are generic systems that enable management of personal information (for instance spatial oriented systems). However, only few studies of user preferences for specific or generic support of activities exist.

One of the conclusions of the study by Nardi and Johnson is that it is extremely difficult to develop software that supports all aspects of tasks well [JN96]. They suggest, "an alternative to task-specific programs is a modular collection of independent interoperable services supporting small subtasks" [NJ94]. A set of small interoperable tools instead of task-specific applications could be advantageous for the support of user tasks. The suggestion is that with small services the user can find and combine the once that fit the task at hand. The highly individual strategies that have been observed in personal information management studies would also support this conclusion. Finding a task-specific application that directly maps to individual work practices is not likely. It must also be taken into account that many factors determine the preference for one type of system over the other. Such factors include usability, user skills, frequency of use, available time, teamwork, and company policy. Therefore the findings in the study by Nardi and Johnson naturally depend such situations of use. If a user carries out a task infrequently this may influence the preference for the type of tool that supports the task.

In addition to this conclusion we note that an important aspect not addressed by Nardi and Johnson is the ability to support changes in tasks over time, which is highly relevant to the support of personal information management. In that case the task-specific applications has to be changed. This depends on the customization possible in the specific applications, which often is not possible. Some software applications allow the user to tailor the application to some degree, for instance in terms of an application scripting language or by installing application plug-ins. Such facilities are often non-trivial and thus targeted at advanced users rather than the average user of the system. Nevertheless these observations support further the conclusion of the study that a modular collection of

interoperable tools could be advantageous in the support of task dynamics and change over time.

5.3.4 Decomposing Available Functionality

Making available a collection of fine granularity interoperable tools can be considered a decomposition of the available functionality in existing systems. Such a decomposed user interface providing fine granularity services across a system has received some researched attention, for instance CyberDesk [DAW98a] [DAW98a], Views [Pem92], Humane interfaces [Ras00] and Services [App01] approaches.

In these approaches modular service functionality is made available across an entire system. Pemberton [Pem92] describes the "Views" project that addressed the problem of "barrier" introduced by applications, and a pilot prototype of the "Views" approach was implemented in the project. Another approach was researched in the CyberDesk project [DAW98a] [DAW98a], and other research by [NMW98], and [PK97]. The approach is to bridge applications by making available commands across the applications in the system. An information entity of a certain type (e.g. text) inside one application could potentially be handled by a number of other applications in the system. The technique is that applications announce their capabilities – the information types can operate on – to the rest of the system. The announced functionality is therefore made available in applications and the user can apply the functionality in a suitable context. In a sense the underlying idea is similar to the Unix commands and pipelining feature discussed in Section 5.3.2. The Apple Mac OS X also has a so-called services architecture [App01] that allow applications to announce their functionality as services – the input that they are capable of handling – to the rest of the system. These services are made available through a specific "Services" menu in all applications that support the services architecture. The user can add his own services through the use of scripting language.

A simple example of such functionality is the user marking a piece of text in an application. The piece of text includes an address and a phone number. The user then invokes a command to add the address and phone number to his address book, or lookup the address on a map.

These approaches aim to glue applications together, and in a sense simulating a collection of smaller granularity tools as discussed in the previous section. Additionally to address the problems that modes introduce in application-oriented user interfaces. Nevertheless, applications are still the point of access to information. Raskin [Ras00] has suggested a design of a generic system entirely without applications as such. In his system the available functionality is decomposed, leading to a set of smaller granularity functionality. The aim is a greater level of consistency, since each function or command is available once and applies everywhere in the system and is operated the same way across the entire system. The aim is also to ease learning and facilitate habit formation, as the commands in the system can be applied for text everywhere in the system, and the way the command is invoked is done in the same way everywhere in the system.

Besides being beneficial from a strict usability point of view, such approach could provide additional support of personal information management in terms of higher level of flexibility in the organization and use of personal information. The information organization and use in the different contexts relevant to activities.

5.3.5 Noun-Verb versus Verb-Noun Type Interaction

Another issue that has implications for personal information management is the type of interaction in the user interface. That is, whether the user interfaces uses noun-verb-based interaction or verb-noun-type interaction. In this context the noun is an information item in the user interface and the verb is as an action that can be applied to that object. Systems may use one or both of these types of interaction. For instance changing the typeface for a paragraph in a document can be done both ways:

- Verb-noun approach: Choose the change font command (the verb) and then apply it to the paragraph (the noun).
- Noun-verb approach: Choose the paragraph (the noun) and then apply the change font command (the verb)

In personal information management where different types of information is used together and interrelated the verb-noun approach would seem inappropriate. This type of interaction introduces problems such as different modes of

operation, and most user interface guidelines recommend noun-verb based interaction [Ras00]. The approach requires the user to choose a specific function or application first and then use and manage information relating to that function or application. In personal information management the primary interest would appear to be the information as such (noun) and the organization and use of that information. The noun-verb approach would require the user to choose the information first and available commands to that information in order to use or organize the information.

Different approaches discussed are also different in terms of the type of interaction. Lifestreams [FFG96b] and spatial-oriented systems like PAD++ [BH94] use noun-verb based interaction. First the information is chosen from a stream or the spatial environment, and then commands are applied to the information. Approaches such as specific personal information management software are typically verb-noun oriented as the specific area (e.g. calendar or addresses) is accessed and then the specific information (noun) is located and used within that area (verb).

5.4 Summary

In the previous chapters the variety of aspects in personal information management was discussed (Chapter 2) including cognitive issue (Chapter 3). That is, the cognitive capabilities that must be considered in order to provide support of personal information management. Existing approaches to support and facilitate personal information management has been discussed (Chapter 4) including the levels of support for the various aspects identified. The discussion has provided critique of existing approaches in terms of supporting different aspects of personal information management. Some of these issues include:

- Flexibility in structuring and organization of information
- Supporting unstructured, semi-structured, and formally structured information
- Different classes of information or levels of interaction: ephemeral, working, and archived information

- Reminding
- Finding and retrieval of information
- Views of the same information in suitable context
- Flexibility in information types used together

This chapter has extended the discussion of support of personal information management with focus on user interface and interaction specific issues. This means issues that have implications for the support of personal information management. The discussion has included user interface and interaction oriented problems that have been identified in studies of human computer interaction and personal information management.

The problems indicate that existing approaches are insufficient in supporting the issues involved in personal information management in terms of information organized in task-specific and compartmentalized ways. It also indicates that an approach where the organization of information and available functionally is unified across the variety of information types is relevant to consider as it may provide a higher level of support of personal information management.

Chapter 6

Models

Personal information management is a complex area. It involves different types of information used at different levels of interaction for a number of different purposes. Also the individual strategies used in order to manage personal information are highly varying. This means that many aspects must be taken into account in systems, which aims to provide suitable support of personal information management. In the previous chapters it was discussed how different approaches have been tried out in the last couple of decades, both in the research and commercial space.

A challenge for interactive personal information management systems is to find suitable models on which such systems can be based. The models must be capable of capturing the aspects of personal information, the ways in which users interact with and manage their information, and the way in which it is stored and organized and later retrieved. Thus an aim is to capture the set of fundamental issues relevant to personal information management.

This chapter discusses such models at three different levels, that is, models on which personal information management systems may be based. This includes the conceptual model of personal information management, the interaction model, and the data model on which a system is based.

These discussions lead to an attribute-oriented model based on findings in studies of personal information management and different approaches to per-

sonal information management, which aim to address the issues identified in the previous chapters. The purpose of the model is to serve as a foundation for interaction personal information management systems, which provide support for the variety of and disparate information used in personal information management. Finally the implications of the model in terms of supporting personal information management are discussed.

6.1 Models at Different Levels

Providing support for of personal information in interactive systems involves issues at different levels, including:

- Conceptual model – the underlying understanding of issues involved in personal information management on which a system is based
- Interactive model – the way in which users interact with and navigate the information available in a system
- Data model – how the actual information is represented and structured in a system

Each of these models has implications for the level of support obtained in personal information management in interactive systems. Also there is an interrelation of the models, meaning that for instance the data model have implications for the interactive model used, and the conceptual model have implications for the data model used.

6.1.1 Conceptual Models

First and foremost at the conceptual level is an understanding of the underlying complex issues involved in personal information management are paramount. This includes the variety of personal information and the variety of ways in which it is dealt with in order to support daily activities. It must take into account the human memory and specifically the way in which humans recall information. This also includes the varying level of structure of the information

and the highly individual ways in which personal information management is performed. These issues have been discussed in Chapter 2 and Chapter 3.

Existing systems have addressed these issues in different ways and emphasized different aspects of the complex set of issues involved in personal information management, as discussed in Chapter 4. Some approaches have primarily addressed personal information management as a matter of archiving information, and with less consideration to the levels of information that includes ephemeral and working information [BN95]. Thus the complexity of personal information management is addressed in different ways in the existing systems. Typically the systems under consideration have only addressed a subset of the issues mentioned.

The systems under consideration are built on different conceptual models of personal information management. This means the understanding of the aspects personal information management on which the model is based. When personal information management is understood as a matter of providing support for archived information the model falls short in terms of addressing the complexity of issues involved.

Some systems tend to focus on specific tasks, whereas others try to provide a generic framework for information management. The task-specific approaches are appropriate where user tasks map easily to the tasks that are supported in the system. However, the studies of personal information management have shown that humans typically create coping strategies or adapt their tasks or strategies to fit the available system, in order to deal with the lack of support of tasks they need to perform. For instance by using the systems in unintended ways [ABC98].

As discussed in Chapter 4 existing systems are based on different conceptual models. To mention a few examples desktop-based systems [JRV+89] are based on a metaphorical approach using documents, a filing cabinet, and applications. Lifestreams [FFG96b] uses a diary metaphor with time-based streams of documents. Systems such as PAD++ [BH94] utilize a spatial metaphor for the organization of documents, images, et cetera. Each of these approaches thus have a set of underlying assumptions about the issues involved in personal information management which is reflected in the conceptual model of the system. This has implications for the level of support of personal information manage-

ment obtained in the systems. Again it should be kept in mind that some of these systems were not created to support personal information management as such originally.

6.1.2 Interaction Models

Another level that is relevant to consider in terms of the support of personal information management is the interaction model. That is, the way in which users may interact with and navigate the information in a system.

Existing systems discussed previously also have different interaction models. Some projects have tried out different ways of visualizing information in order to provide a better overview of the information contained in a system. Some models focus on a specific property of the information on which the conceptual model is based, for instance information organized on a temporal basis.

The systems mentioned above provide examples of different interaction models. Desktop systems [JRV+89] provide a set of applications and the desktop, which provides ways to navigate and manage information. In Lifestreams the user navigates streams of documents with a small set of operations [Fre97]. In spatial oriented systems such as PAD++ [BH94] the user navigates a virtual space in order to manage and retrieve information. Each of these interaction approaches has advantages and disadvantages in terms of providing support for personal information management, as discussed in Chapter 4.

One issue is whether the approach is task-specific or generic in terms of personal information management support. As an example desktop-based systems have applications that are general such as word processing, spreadsheets, and drawing. These typically do not map to the tasks that people perform, such as, writing a letter, a report, or a note, which may involve other information and functionality as well. Thus the tasks performed by humans are more complex than simply word processing. The writing of a report may need the incorporating of information of many different types and from other sources. Although current software applications are powerful in terms of available commands they may lack the support of the information and functionality needed for any given task [Nor98].

Especially for personal information management this is a crucial issue. It has been argued that the variety of information involved in personal information management is used in different ways across information types and functionality and combined in flexible ways. Thus task-specific conceptual models may seem less appropriate as the underlying model for personal information management.

6.1.3 Data Models

Finally the data model on which a system is based is crucial for the support of personal information management. The data model includes the schemes and methods by which information is stored and organized. Also the data models in existing systems are quite different and naturally have implications for the way in which systems based upon the model can provide support for personal information management.

The data model must support a wide variety of information types, activities, and organization methods. Based on the findings in studies of personal information management especially support for unstructured and semi-structured information is required to support organization and management of personal information. Additionally flexibility in the model is needed to support information structures that evolve over time.

Traditional systems have a tendency to provide fixed information schemes. For instance specific personal information management software applications provide specific types of information, such as calendar items, contacts, and to-do items each with predefined and fixed information structures. Users must fit their information into these fixed and predefined schemes, which leaves little flexibility. Such approaches may also prevent evolvable information structures.

Database systems also provide static structures in the sense that changes to underlying tables and data structures is done by a developer of the system rather than an end-user of the system. Thus changes to underlying program code is needed in order to support new or changing structures and schemes, which provides little flexibility in terms of information organization. Structures that are static do not provide support for the dynamic, changing, and evolving information (types) typically relevant in personal information management. A higher level of flexibility in the model is needed in order to provide better support

of personal information management.

In the previous chapters it was discussed how hierarchical information structures is a common way of supporting information management in interactive system. For instance it is common in desktop-based systems, as well as in email, bookmark, and document organization systems. There is a tendency to require information to be organized in a hierarchical structure whether the information naturally fit into a hierarchical organization or not. Thereby humans have to adapt their practices to the structures enforced by the systems. For instance ephemeral and working information mentioned does not need such an information structure. Hierarchical information organization does not support these aspects of personal information management in a suitable way. This has been observed in studies where subjects tended to come up with various strategies in order to cope with the lack of support of some aspects of personal information management [BN95]. Additionally such systems typically only allow the information to be put in exactly one place even though it may belong in several places in the information structure.

Hypermedia based systems have provided more flexibility in terms of information organization. Such systems typically provide a graph of interconnected nodes with no root or information structure provided a priori. Thus it provides a higher level of flexibility in that any node in the system can reference any other node. This also means that information structure may evolve over time, as new information may be linked into the existing information, and existing nodes can be changed to reference new information and new structures. However, multiple organizations of the same information are not supported directly – only in the sense that each node can have multiple links. This means an information entity may have references to multiple information entities.

6.1.4 Summary of Issues

In order for interactive systems to support and facilitate personal information management modeling at several levels are relevant to consider, as illustrated in Figure 6.1. At the conceptual level is the understanding of issues involved, at the interaction level is the interaction and navigation of information and at the data level is the representation and structure of information.

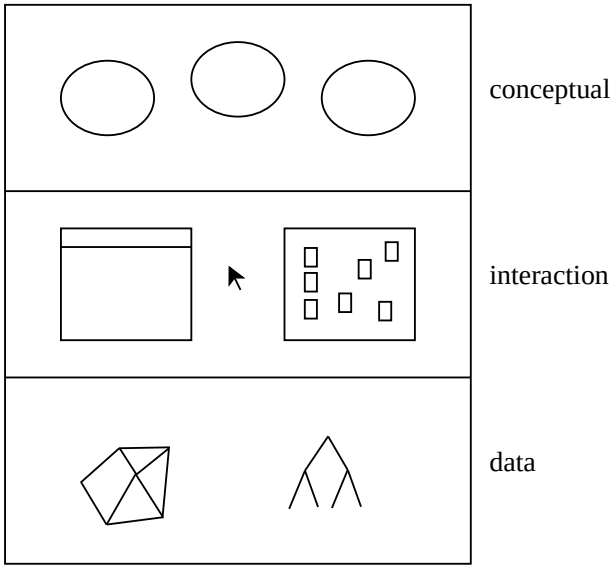

Figure 6.1: Different levels of issues in support of personal information management, conceptual level, interaction level, and data level

One of the most prominent problems in existing systems is the data models on which these systems are built. As mentioned above the hierarchical information structure is widespread as the basis for personal information management systems such as email, bookmark, document management, and general file systems. Traditional systems tend to use such fixed data models having some kind of predefined information structure (such as hierarchical structure). Such data model approaches are inflexible in terms of complying the specifics needs in support of personal information management. They may serve different needs and enable other types of applications, but they have shortcomings in terms of supporting some aspects of personal information management in an efficient way. Personal information management system needs to support a high level of flexibility in terms of how the information in structured in the system, as semi-structured and unstructured information must be available as well. Some personal information is created and used on an ad-hoc basis, which requires more flexibility than

existing schemes may provide in an efficient way.

Pre-defined information structures and schemes can be observed in systems with fixed applications, which maps to tasks domains. These systems often have predefined data schemes, which means that they only accept fixed data elements. The immediate implication of such fixed system and model is that the user of the system has to adapt his personal work task or information need to that of the system or map it to the structures supported by the system. More flexibility is needed due to changes over time, the information usage, or the context of use evolving over time.

The way in which systems are designed and engineered today makes them uneasy to adapt to new or changing demands. Changing requirements means changing the underlying structure of the systems. This implies changes to the schemes, which is sometimes not possible at run-time. It means updating data schemes and compiling program code in order to change a system to support different information structures and work tasks, which is not easily done. This is especially difficult if not impossible for the end-user of the system.

Some of issues mentioned above can be summarized as follows:

- Observing the complexity of issues in personal information management.
- Supporting unstructured and semi-structured information and ephemeral and working information.
- Hierarchical organization of information leaves little flexibility, is fairly static, and allows only one place to put an information items.
- Need to keep a scheme, but the use and organization of information changes over time.
- Naming and identification of information items is not always easy to do, or even relevant to do.
- Adapting systems to changing requirements and new information types and structures is not easily done in current systems.

6.2 The NEXUS Model

The approach researched in this project involves an attribute-oriented model on which personal information management systems can be based. The model consists of four key elements:

- Information entity – the basic unit holding information
- Attribute – an information entity related to another information entity containing information (metadata) about the information entity
- View – a collection of information entities, that is, a subset of the information entities contained in the system
- Action – functionality which may be applied to entities, attributes, and views (or any combinations of these)

The external representation of the information in the model is the *repository*, which is the entire collection of information entities (and attributes, views, and actions).

Besides the four key elements the model have a few additional elements:

- Template – an information entity with a set of default attributes associated
- Reference – an attribute which contains a reference to another information entity
- Rule – a definition of an action to perform based on a set of conditions

6.2.1 Entities

Information entities are the lowest granularity information unit in the model. An information entity holds a body containing the information it carries. A simple information entity can for instance hold a text string as shown below.

```
("This is an example information entity")
```

The information in the model is described as a collection of information entities. That is, the external representation of the information –the repository– is the entire collection of information entities. As attributes, views, and actions are special-case entities they are also included in the collection of entities in the repository.

The repository can also be described as a *multi-set* of entities, in that, two (or more) entities can be identical (semantically equal). However, all entities may have a unique reference, and be uniquely referenced in the collection.

As seen above an entity may represent a simple piece of text. Other examples of simple information entities include:

- Call Michael about meeting
- Birthday party the 5th
- Remember to buy milk

The point is that the entity is information at a fine level granularity, and that they may or may not have an internal structure as such. Individuals can use and manage many such information entities on a daily basis depending on personal strategy and style.

Information entities are not necessarily personal or private, but may include information that is of personal interest. For instance an entity could hold a reference to a document available on the World Wide Web that contains information of personal interest.

6.2.2 Attributes

An attribute is an information entity that describes or provides information about another information entity. Each information entity has a set of attributes associated. Thus attributes represent meta-data, that is, data that describes other data (information entities). However, the associated set of attributes may be empty.

Attributes themselves can be described as name/value pairs. Thus an entity can have any number of such name/value pairs associated. Including the infor-

mation entity in the description means that attributes can also be describe in the form (entity, name/value).

As an example an entity representing a contact person, may have a number of attributes (name/value pairs) associated describing the contact – the person:

```
("e-mail address" / "john@doe")
("address" / "First Street 1st")
("zip" / "12345")
```

Any number of attributes can be associated to entities, thus the approach can be described as a set associated with the entity. Elements can be added and removed from the set as needed.

The example provided above also illustrates that the attributes form statements about an information entity. In the example above there is an information entity representing a contact, which has an associated attribute – the name/value pair e-mail address, and the content of the e-mail address.

6.2.3 View Collections

Information entities can organized into any number of groups – collections. Such a collection of information entities in the model can be organized in different ways based on the attributes associated. Based on criteria of selection the information collection can be organized into a sub-collection of entities, that is, a subset of the entire multi-set of entities.

A collection of information entities is simply a subset of the entire set of information entities (in the repository). A collection is also referred to as a *view*, as it may provide different views of the information entities in the repository.

```
view = {entity 1, entity 2, ... , entity n}
```

As mentioned above the attributes associated with information entities can be added and removed from the set associated with an information entity. Therefore it is the associated attributes that constitute the structures of information entities in the repository, in terms of any number of defined collections.

For instance a collection or view that should represent "e-mail addresses" from the repository could be defined by the criterion including all entities that have an "e-mail address" attribute associated. Thereby the collection will hold a set of information entities that all have an e-mail address, and thereby establish a view of email addresses from the repository.

The selection criterion may be based on any attribute (name and value) or combination of attributes (name and value). This means views can be established on the basis of information entities having several attributes associated, also that the value of attributes must meet a certain criteria. For instance a view of Danish e-mail addresses could be defined so that it would only include e-mail addresses (with values) ending with ".dk".

This approach enables the creation of any number of views of the information entities in a repository, and for the information entities to be included in multiple views.

6.2.4 Actions

The information entities and attributes represent the information that is in the repository (the entire set of information entities). The collections define different ways to view the information entities and attributes in the repository. The next element is actions that represent functionality that may be applied to information entities, attributes, and collections of information entities (or any combinations of these) in the repository.

Core actions include the creation and deletion of information entities and attributes, and the creating or altering of collection criteria, in order to create or modify views of information.

Further actions can be defined to allow a specific functionality to apply to certain information entities, attributes, or collections by some criterion. The approach described here does not enforce information to be of any particular type, however, information entities having a given set of attributes associated may be interpreted as having a type.

For instance an information entity representing an e-mail typically could have the following attributes (name/value pair) associated:

```
(sender / ...)
(date / ...)
(to / ...)
(cc / ...)
(subject/ ...)
(body / ...)
```

Actions that apply for e-mail information may be defined so that they apply to information entities having some or all of the above listed attributes associated. Such an action could be to send an e-mail to the address or addresses provided by attributes providing the to and cc of the e-mail. The action could automatically associate sender and date attributes to the set of attributes. In addition it would for instance add an attribute indicating that the e-mail has be sent, so that the information entity would be included in a collection defining sent e-mails.

Thus actions can be defined so that they apply to information entities, attributes, and views based on a relevant criterion. Actions may include functionality, that is, some manipulation of information entities, attributes or collections of entities. As there are no types of information as such actions apply across the entire set of information in the system as long as the defined criterion is met.

6.2.5 Templates

Templates are information entities with a number of default attributes associated. Thus a template may hold an entity with or without content with a set of associated attributes. This element allows for the construction of frequently used information entities and structures.

For example an information entity holding a reminder (a to-do item) could have a template with the following attributes:

```
(urgency / ...)
(priority / ...)
```

A new information entity based on this template would have default associations to attributes `urgency` and `priority`. In the example specific attribute names are associated and the attribute values are left blank. After creation by means of a template it is possible to fill the information entity and attributes in order to make it represent the needed information.

As seen in the example in the previous section an information entity holding an e-mail would also have a set of default attributes. Such an e-mail template could serve the functionality of creating a new e-mail message, and the example action mentioned above would provide the functionality of sending the e-mail.

6.2.6 References

As mentioned previously humans may deal with many information entities on a daily basis. Above information entities were described as fine granularity information units, such as, a contact, an e-mail, a to-do item, et cetera. However, it is also necessary to deal with higher granularity information units, such as, documents, reports, presentations, et cetera. In that case an information entity may simple serve as a proxy for that information, in that it provide a reference to the information that may origin from another source. The information entity could for instance hold a short summary of the information being referenced. The reference itself can be an attribute associated with the information entity.

The information entities can also reference other entities in the collection of entities. A reference is an associated attribute that holds a reference to an information entity. Thus each entity can have any number of references to other entities in the repository. For instance an entity may have the following attribute name/value pair associated:

```
(attribute name / entity 2)
```

In this example `entity 1` has an attribute that holds a reference to `entity 2`. The attribute could describe the kind of reference, or simply indicate that it is a reference or link.

This approach enables the information entities in the repository to have multiple organizations and structures, includes different types of information structures. Also note that entities are not required to have associated attributes with

references to other entities, thus enabling unstructured information entities in the repository as well.

6.2.7 Rules

Finally rules define actions to perform if some criteria are met. Based on the content of the entities and the associated attributes rules can be applied. Actions to perform include:

- Attributes added, removed, and/or changed
- Entities added, removed, and/or changed
- Views added, removed, and/or changed

For instance if an information entity has associated attributes `subject` and `to` it can act as an e-mail as described previously. Rules can be defined for different purposes, such as flagging specific information entities in terms of associating additional attributes. In the case of emails messages with a particular subject or sender could be flagged (attribute added) indicating important.

6.3 Implications for PIM

As discussed in Section 6.1 different models and concepts have been chosen in various personal information management systems. The choice of underlying models has implications for the level of support that can be obtained in the system, or the set of aspects addressed. Especially there are implications for the way in which information entities can be organized in terms of the available elements in the models.

The issue of different data schemes has been emphasized as a problem in terms of supporting the variety of ways in which personal information is organized and managed. Fixed schemes and predefined structures pose problems in terms of supporting the flexibility and variety of information organizations found in personal information management.

A discussion of the implications of the approach presented in the previous section follows. It is emphasized how a set of the issues in personal information management is addressed by this approach in contrast to existing approaches discussed.

6.3.1 Entities and Attributes

Information entities, attributes and collections are the core elements. Information entities are the basic units representing information. These information units may represent simple information elements such as a piece of text, a note, an appointment, a to-do item, et cetera. This approach enables a fine granularity access to information. The information entities are identical in the sense that information entities are not of any type. Information entities simply represent an information unit regardless of the kind of information that it holds.

Different *types* of information are obtained in terms of the sets of attributes associated with them. As seen in the examples in the previous sections this enables information entities to represent e-mails, to-do items, contacts, or any other information. This depends on the attributes associated. This also means that an information entity may represent information of different types at the same time. For instance a bookmark can also serve as a reminder with the proper attributes associated. Any attribute may be associated with any information entity, which provides a high level of flexibility of information use and management.

Information entities may represent personal information or general information. E-mail is typically personal (even though it could have multiple recipients). Whereas a bookmark is a generic entity, that could reference a public information unit. However, attributes associated with such a bookmark could be personal, if they represent personal interests of the particular reference. For instance the context in which this bookmark is found to be useful or of interest to the individual.

6.3.2 Collections and Information Structures

Information entities are organized into collections, which simply represent subsets of the entire set of information entities. The collections are defined to include the information entities where the attributes associated meet a certain criteria. This approach has benefits in terms of providing additional flexibility of information organization compared to existing approaches as the model allows arbitrary relationships among information entities. It can be argued that the model is useful for the support of personal information management as the information organization allows a high level of flexibility in terms of information structures (unstructured, semi-structured, and formally structured).

The model enables different ways of navigating and retrieving information entities available in the entire repository of information entities. Attributes provide the basis for collections and the attributes may form these collections so that they have a particular structure. Different attributes can be associated with the information entities, which may be used to create different views (visualizations) and structures of the subsets of information entities. Such visualizations may enhance the visual perception when reflecting an underlying structure of the information. Different structures include:

- Temporal
- Piles (or lists)
- Hierarchical
- Spatial

This approach enables multiple ways of navigating and finding information. The subsets of information entities can be overlapping, in order for an information entity to appear as part of different information structures, see Figure 6.2. As an example an information entity representing a person might appear in different contexts, e.g. as the sender of an e-mail, in an address book, on a list of participants in a project, et cetera.

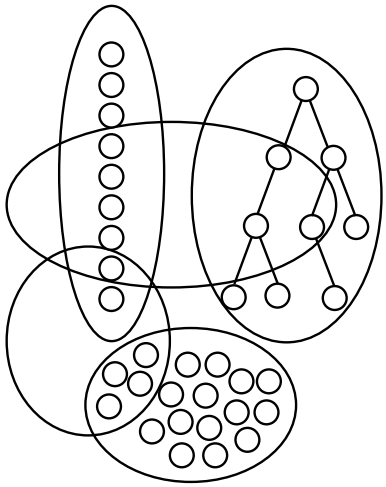

Figure 6.2: Flexibility of information collections
The figure provides an illustration of different types of collections of information entities: linear/temporal structure, hierarchical structure, and semi-structure, and two collections that span across the other collections.
o = information entity, and O = collection / view

Additionally the model does not enforce any specific information structure. The arbitrary relationships among information entities allow these collections to have structures as those mentioned above, as information entities can be grouped and linked in arbitrary ways. This allows information to be organized into loosely structured groups such as a set of information entities having a specific attribute associated (see the e-mail example in Section 6.2.4 above). Another possibility is to have a collection of information entities with temporal attributes and provide a temporal ordered collection of information entities. Such is naturally useful for calendar information and a diary oriented approach to information organization [FFG96b], [Rek99], and piling of information [MSW92]. Spatial organization [BH94] is possible if attributes representing spatial locations is associated. For additionally structured information a collection of information entities may take the form of a hierarchy, in that associated attributes provide information for such a hierarchical information structure.

The highly individual information management practices may be supported in terms of various ways of organizing and managing information. This includes the model providing the ability to support unstructured, semi-structured, and formally structured information. Thus it may provide support of both piling and filing strategies of information management and organization [Mal83]. This also forms a basis on which to provide support of different levels of interaction with information – ephemeral, working and archived information. At least at the level that the various levels of structure that characterize these levels of interactions are supported by the model.

6.3.3 Multiple Organization of Same Information Entity

Information entities may not be easily mapped into an information structure. For instance there may be more than one possible location for an information entity in a hierarchical information structure [Lan88] [Mal83]. An information entity may belong to different categories or classifications as it may be used and referred to in different contexts of use.

The model allows for information entities to by part of different information collections, as illustrated in Figure 6.2 above. Thus there may be multiple views of the information entities (illustrated by the large circles in the figure) and an information entity may appear in more than one. Therefore the model allows multiple organizations of the same information entity, in that it can be part of different collections, and therefore also different information structures.

This also allows for multiple ways in which an information entity may be retrieved in a system built on top of the model. An information entity may be part of different structures that can be used as the basis for retrieval. Also the information entity may have multiple attributes associated providing descriptions of the information may be used in the retrieval process. Potentially there are multiple ways of facilitating retrieval, and thus utilizing more dimensions of human memory in the retrieval process [Lan88].

Finally, similar information structures may be used across different collections of information entities. As seen in some studies [Boa01] [Kwa89] subjects maintain parallel information structures across different types of information (e.g. e-mail and documents) relating them to particular activities or use. As the

model does not impose restrictions in terms of particular information structures it is possible to create such parallel structures. For instance in an organization with one collection for each type of information entity. However, as mentioned above the model also allows different information entities to be organized in the same collections. Using the example of e-mail and documents, meaning that those could be organized into one collection if needed.

6.3.4 Identification of Information Entities

As discussed above the model does not enforce any particular information structure – no structure is specified a priori. The structures are entirely formed by the collections determined by arbitrary attributes associated with the information entities.

Another problem identified in traditional information management systems, such as file systems is the need for explicit naming of information entities. Files in file systems need to have an explicit name when created. In some situations it is difficult to provide a useful name, and humans tend to forget names of files [BN95]. This particular issue was addressed by the Lifestreams approach [FFG96b], in terms of using temporal ordering by default. This model does not enforce naming of information entities either, as the association of attributes to information entities is optional. The set of attributes may even be empty. This means that any attribute can be associated with information entities proving different types of descriptive information about the information entities. This includes temporal information, structural information, identification information, references, et cetera. Thus it is possible to provide an attribute that contains a name if it is suitable for a particular information entity. However, the model does not require explicit naming or identification of information entities

This approach aims at supporting both the piling and filing strategies of personal information management. When filing information naming, identification, and particular structuring of the information is typically necessary. The attributes support the possibility of using such a filing strategy to organize information entities. When it comes to loosely and semi-structured information, such as piling information, it is often done by different means. Information can just be left in a "pile", and retrieved for instance by an implicit temporal attribute.

Thus providing naming and identification of the information entities would be troublesome and unnecessary.

6.3.5 Changes Over Time – Evolving Information Structures

One issue is that information may be used in different contexts meaning that multiple organizations of the same information entity may be relevant, as discussed above. Another issue in personal information management is that information structures tend to evolve over time when the context in which the person information is used changes over time. The consequence is that information may have to be reorganized and restructured in order to reflect new situations of use. The underlying structure may impose difficulties in terms of gradually changing information structure and use over time. Highly structured information may be difficult and time-consuming to change. For instance a reorganization of hierarchical structured information may be extensive as the underlying classification of information entities need to change [Lan88]. This could potentially lead to the reorganization of a considerable part of information entities in the structure. Not all approaches to personal information management address this issue of information accretion and changes over time.

New activities, situations, or context of use may have an effect on information structures, that is, causing them to evolve over time. The attribute-oriented approach may provide some level of support for information structures to evolve. For instance by allowing multiple structures of information entities. This means that the different contexts of use may be supported by association of suitable attributes, as discussed above. However, the use of attributes enables structures to be gradually changed or expanded, as it is possible to change or add attributes over time. This makes it possible to add new collections or modify existing collections of information entities to create new views of information entities. Thereby having them support new contexts of use. By changing attributes associated with an information entity it may include the entity in or exclude the entity from a collection. Thus a collection may change over time, as information is changed or new information received.

Nevertheless, the problems of evolving structures are also influenced by the choice of information structure. If the attributes are used to construct a hier-

archical information structure it does not change the fact that reorganization of a hierarchical structure can be cumbersome if extensive changes to the underlying classification of information needs to be carried out. Unless actual implementations of the model provide facilities that facilitate the reorganization of information structures at the level of interaction the model is only advantageous because it allows multiple organizations of information. Nevertheless, as studies of personal information management indicate, hierarchical organization of information is mostly used for long-term archived information, making it less useful for frequently changing information or situations of use.

6.3.6 Reminding

It has been emphasized that reminding is a key element of personal information management. However, the model does not have any inherent support of reminding. Even though there is not a specific element in the model addressing the issue of reminding there are different ways in which personal information management systems created on top of the model may support reminding.

As pointed out in studies subjects often use various strategies in order to have some support of reminding in existing systems. For instance by leaving information entities in particular locations in which they will be frequently seen, such as e-mail inbox or in some spatial arrangement. As such it could be argued that information entities may simply be left in a suitable collection that is frequently seen. That would provide at least at the level of support as in existing personal information management systems. That is, a similar method of reminding is possible with this approach.

In addition any information entity can have attributes associated, which provide an indication of a reminder. Thus an information entity may have arbitrary attributes used as part of a reminding scheme. Such associated attributes could include:

- Urgency – how urgent is this reminder on some scale (e.g. temporal)
- Importance – how important is this reminder on some scale (e.g. priority)

In effect any information entity may serve the purpose of reminding, and may have attributes associated informing of the kind of reminder.

6.4 Summary

This chapter has discussed models of interactive personal information management systems on three different levels, the conceptual level, the interaction level, and the data level.

The conceptual level includes the understanding of the issues involved in personal information management (discussed in Chapter 2 and Chapter 3). The interaction level includes the ways in which users interact with and navigate the information available in a system (discussed in Chapter 5). The data level includes the actual information representation and structure in a system. Different data level approaches in existing systems have been discussed briefly.

This lead to a presentation and discussion of a unified attribute-oriented model for personal information management systems. The model includes four key elements, information entities (the units of information), attributes (information associated with information entities), views (a subset – collection – of information entities), and actions (functionality applied to entities, attributes, and views). Examples of these elements were provided.

Moreover the implications of the unified attribute-oriented approach for personal information management have been discussed. This includes a discussion from the perspective of the conceptual understanding of personal information management discussed in Chapter 2 and Chapter 3. Especially there are a number of implications for the way in which information may be structured and organized in collections using this approach. The discussed involved issues that have been identified as essential for personal information management, such as flexibility of information structures, enabling multiple structures, evolvable structures. Finally the way in which reminding could be supported by this model was briefly discussed.

Chapter 7

Research Prototype

This chapter discusses a research prototype implementation of a personal information management system. The prototype system – NEXUS – is an application of the attribute-oriented model described in Section 6.2 and the system also provides a user interface that aims to provide a unified user interface and interaction in the system as discussed in Chapter 5.

This chapter provides an overview of the research prototype implementation and a description of the system architecture and platform used. This includes the repository storage system and a framework for mapping existing information sources to the system. Additionally the user interface of the system is described including the implementation of entities, attributes, views, and actions and applications of the system for personal information management. Finally a set of limitations of the research prototype implementation is described.

The research prototype represents a pilot implementation at a proof-of-concept level. The purpose is to enable initial experiments to be carried out, thus allowing initial exploration of applying the attribute-oriented model combined with a unified user interface for personal information management.

7.1 System Overview

The research prototype is an implementation of a personal information management system for management of common electronic information. The system is implemented as a software application that runs on standard desktop or laptop computer systems. However, an implementation of the underlying model and a tailor-made user interface for mobile information appliances should be possible. In terms of carrying out initial experiments the current version is sufficient.

7.1.1 Information

The system enables the user to keep, manage, and use typical personal information, such as emails, calendar items, to-do items, documents, bookmarks, et cetera. All information is available through the common repository and likewise all information is accessed through a single unified user interface.

The system is based on the attribute-oriented model presented in the previous chapter, thus all information items kept and managed in the system are represented by information entities and attributes. Each information entity can have any number of attributes associated. The information is organized into any number of collections (views) by means of the attributes associated with the information entities. The types of information that has been used with the prototype includes:

- E-mail – messaging
- Calendar items – organizer and diary
- To-do items – reminding
- Text items – memos and documents
- Bookmarks and references
- Music – sound tracks

The list provides and illustration of how the research prototype applies the model in order to support different information types in various domains to

support personal information management. The mapping of these different types of information to the model is not defined by the system. The model provides a flexible way of representing information entities with associated attributes, thus it is possible to do this in different ways. The mappings used for e-mail entities and sound entities are discussed in Section 7.2.5.

7.1.2 Templates

The attribute-oriented approach enables the research prototype to store arbitrary types of information, meaning that the support of information entities is not limited to those mentioned above. The system allows the user to create new information entities, and associate arbitrary attributes with the information entities. Templates can be created in order to facilitate the process of creating new information entities. An information template is simply implemented as an information entity with a number of attributes associated. This means when creating an information entity based on a template a new information entity is simply created with a list of default associations of attributes.

7.1.3 Views

Similar the information is organized into different collections that allow the user to view the information entities available in the repository. These collections are specified to include information entities having specific attributes associated. Criteria for the attributes may also be specified. For instance a view could include all information entities in the repository having a temporal attribute associated, thus creating a temporal ordered view of a subset of information entities in the repository. Individual views may be sorted by the different attributes included in the view.

A set of simple default *views* have been created to enable the user to view and interact with the types of information listed above, including:

- Inbox – e-mail received
- Sent – e-mails that have been sent
- Contacts – address book
- Calendar week overview – calendar items with date this week
- To-do – reminding items
- Bookmarks – references to webpages
- Documents – memos and documents
- Music – sound tracks

Additionally there are views that include information across different types. For instance a sample project view that includes calendar items, to-do items, e-mails, and documents, that are all associated with the sample project.

The research prototype enables existing views to be changed and new views to be created. For instance the default view of music simply includes all sound tracks available from the repository (with title, artist, album, et cetera). As mentioned above views may be sorted by different attributes. However, the user could also create different views of sound tracks by artist, genre, or other criteria.

It is possible for the user to create views of new types of information that does not already have suitable views. For instance a user could create information entities representing *facts* and create suitable views to view the facts available in the repository. The user could for instance create a scheme for organizing the facts, by association of suitable attributes.

Such information created by the user could also be used in existing views by associating attributes to the information entities thereby including it in an existing view. Generally any information entity in the repository can have any

number of annotations by associating of descriptive attributes. The following is an example of such an associated name/value pair:

```
("Summary" / "Round menus are more efficient than menu lists")
```

In a sense all associated attributes serve as annotations to the information entities. This attribute-oriented approach also allows information entities to be related in arbitrary ways, as attributes may include a link to another information entity. This enables information entities to be linked in lists, structured in a hierarchy, or a graph, or any other structure that is suitable. It is noticed that the same information entity may be part of multiple such structures.

7.2 System Architecture

The research prototype is a combination of the attribute-oriented data model and a user interface providing unified interaction with information in the system.

7.2.1 Implementation Platform and Programming Libraries

The research prototype is implemented in the Java programming language [AG96] and runs on top of the Java platform version 1.3 available for typical desktop operating systems. The choice of Java as the development and experimental platform was primarily due to the wide variety of APIs and programming libraries available for the platform. As the prototype needed to have interfaces to different types of existing systems it was necessary to use a platform providing libraries for existing systems.

However, the use of the Java technology is not crucial for the realization of the concepts and approach in an implementation of a research prototype. However, it was considered as an appropriate platform to create a pilot implementation of the concepts with manageable development effort. This was especially important in terms of bridging the system to existing services and technologies. The various libraries allowed the development of connectors to different systems without extensive development effort.

The system user interface is implemented using the Java Swing components. This allows the implementation of a standard desktop application graphical user interface. Other libraries used in the implementation includes:

- File input/output APIs part of the standard Java libraries (java.io packages)
- Networking APIs part of the standard Java libraries (java.net packages), including HTTP transfer protocols
- JavaMail API version 1.3, which support the retrieval and sending of emails, supporting standards POP3, IMAP, and SMTP.
- Images using standard libraries (this includes support for JPEG and PNG standards)
- Sound standard libraries (javax.sound packages) and JLayer MP3 API (version 0.08)
- MP3 ID tag library to read and modify attributes of sound tracks in the MP3 format

The use of these APIs enables integration of existing information and systems with less development efforts.

The prototype system was built and used on a 500MHz PC with 256MB RAM memory, 12GB hard drive and a 15-inch monitor. Especially the issue of response times [Mil68] is critical in the user interface. The research prototype have a reasonable performance as the hardware configuration mentioned above allow typically response times below 100–200 milliseconds. This is the case even with the performance intensive tasks, such as changing views and sorting views by a different attribute.

7.2.2 Architecture Overview

An overview of the system architecture of the research prototype is shown in Figure 7.1 below, presenting the main components of the system. The components of the system are implemented in a package hierarchy *nexus* containing

main components. Components and interfaces related to entities are located in the *nexus.entity* package, components and interfaces for attributes are in the *nexus.attribute* package, and view components and interfaces are located in the *nexus.view* package.

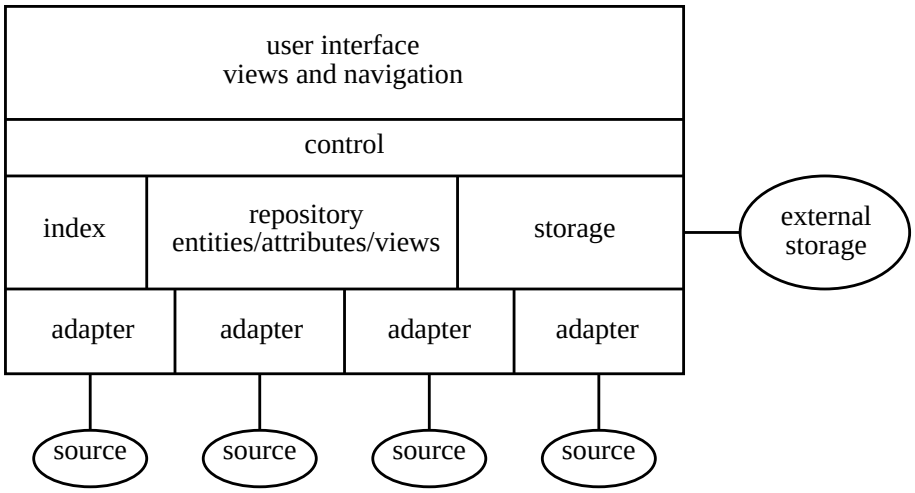

Figure 7.1: NEXUS system architecture overview

The main component is the *repository*, which holds the information entities, attributes, and views available to the user of the system. The actual information may be stored in different locations. The repository itself has an associated *external storage* where the actual information is stored. Access to the external storage is provided through the *storage* interface component. The type of external storage is not essential as the access is provided through the storage interface

The information in the repository is indexed in order to enable efficient retrieval of information from the repository and efficient construction of views of information entities. The *indexing* component is responsible for indexing of information entities, attributes, and views.

The information available in the repository may be stored in the associated storage, but information from different sources may also be linked into the repos-

itory. These sources can be both locally available information as well as resources available through a network. The *adapters* provide the interface to these external information sources. Any number of sources may be connected to the repository through adapters. An example source could be an email server or a local file directory with documents.

Finally there is the *user interface* component on top of the system, which provides the different views of information entities and enables the user to navigate the information, attributes, views, and issue available commands (actions). Interaction with the system is handled by the *control* component that serves as a mediator between the user and the system, and linking the user interface to the underlying system.

7.2.3 Repository Storage and Adapter

The repository is the central component in the architecture, as it stores and coordinates all information available in the system. All components and interfaces related to the repository are located in the *nexus.repository* package. This includes the actual information entities, attributes, and views available. Each information entity in the repository has a unique identification that is used to reference and identify information entities. This identification is used internally in the system only and it not available to the user of the system.

The information in the storage is stored in an external storage, through the storage component. The component abstracts the type of storage from the repository, meaning that it use different types of storage mechanism. This includes different types of database systems or file system. In the actual research prototype a file system is used as the external storage mechanism. The interface of the storage component simply provides the ability to store from the repository and retrieve data into the repository. The component transforms the data into a suitable representation in the external storage. In the prototype implementation a trivial approach is used, that is, data is stored is simply mapped into files that are stored in a file system directory structure. Each unit represents an information entity and is mapped into a file. The file contains a list of name/value pairs representing the attributes associated with the information entity. Similarly the views of information entities defined in the system are stored into the external

storage. A simple example of an information entity representing a bookmark with associated attributes is shown below:

```
Title=CTI homepage
Date=01-01-2004
URL=http://www.cti.dtu.dk/
Project=CTI
Keyword=ICT research
```

The performance of this approach is not optimal, but the prototype performs sufficiently to allow a trivial implementation of the approach. Moreover the abstraction of the functionality into the storage component enables the external storage to be replaced without altering the rest of the components in the system. A system used on a wider scale could thus optimize this part of the system. For instance by storing the information into a database.

7.2.4 Adapter Components

Information entities as the one in the example mentioned above is stored into the repository in the external storage. However, also information from external sources can be accessed through the repository. The connection of external sources is accomplished with adapter components and interfaces (located in the *nexus.repository.adapter* package), which are responsible for the mapping of the information from an external system into a suitable representation that is useful in the repository. This includes mapping the external information entities to an information entity in the repository. This means that each information entity in external storage is represented by an information entity in the repository. Additionally attributes that can be extracted from information entities from an external source are mapped to attributes in the repository. Thereby information from external sources has a representation in the repository, as they were stored directly into the repository.

If additional attributes are assigned by a user they are either stored in the repository or the external storage. This depends on the implementation of the adapter, that is, whether the information source allows additional information to be assigned. Typically additional attributes are stored in the repository. This

means that attributes assigned to an information entity may reside either in the information source or in the repository. The unique identifier mentioned previously couples the two sets of attributes for information entities so that the attributes appear as one set of attributes associated with an information entity.

Thus the adapter components provide a bridge to existing information sources. The open architecture of the system implements a plug-in feature enabling new adapters to other information source be added to the system by means of a standard interface to the repository component. Specific adapters are implemented in order to handle different types of existing information sources. The adapters define the ways to interact with an information source (retrieve and store information) and how to extract defined attributes from information entities provided from the information source. The approach allows the entities to exist in the repository of entities in the system, and for the user to have access to these like any other information entity in the system.

The information sources can be both locally available information sources as well as information available through a network. At first this is not important for the repository as such as an information entity itself is the same whether it origins from a networked or non-networked source. This means that there is no difference from a user point of view whether information in the repository has a networked source or not. However, as network failures are inevitable the specific adapters for networked information sources must implement suitable error handling in case of network failures [LR02]. This is relevant both in the retrieval from and storage into an external information source.

7.2.5 Information Sources

The research prototype implements a number of adapters to existing information sources by means of the programming libraries mentioned previously. The experimental prototype adapters implemented in the research prototype include:

- IMAP – access to e-mail from IMAP e-mail account
- File system – access to files in the local file system
- HTTP – networked access to files via the HTTP protocol

- MP3 collection – access to sound files (in MP3 format) through a file system
- Picture collection – access to image files (JPEG and PNG) through file systems

The *IMAP adapter* enables the system to retrieve e-mails from and store e-mails into an e-mail account on an IMAP server. The adapter uses a mail API to implements functionality to connect to the server and retrieve e-mails from the server. Each e-mail is mapped to an information entity and metadata attributes such as From, To, Cc, Subject, Sender, and Date are extracted from the e-mails and mapped to attributes in the repository. This means that e-mails from an e-mail account can be accessed through the repository. As mentioned above a default view of the inbox provide a way to view the e-mails available in the e-mail inbox. Additional attributes may be associated with the e-mail information entities, if the user wants to categorize these or provide additional information. These are stored in the repository, and therefore not dealt with by the adapter.

The *file system* adapter maps a directory of files to the repository. Each file is mapped to an information entity, and the adapter provides a mapping of filename, date, size, and other file attributes to the attributes in the repository. In a similar way the *HTTP adapter* map files to information entities and extract file attributes and map these to repository attributes. The only difference is that the latter adapter provides information from networked sources. Therefore the adapter includes error handling in case of network failures, such as the system being unable to retrieve information.

In the *MP3 adapter* the information source is a local directory containing files in the MP3 sound format. The adapter could be modified to access files from a networked resource by means of the previously mentioned adapter. Music entities have several attributes that are extracted from the files and mapped to attributes in the repository. These attributes are stored in so-called MP3 tags, and include: song title, artist name, album title, length of track, genre, et cetera. These attributes can be used to form different ways of organizing the music tracks, that is, views representing play lists enabling the user to play an album, all tracks by an artist, or a particular genre, et cetera [Ros00].

The *picture adapter* also uses a local directory as the information source. In a similar way it maps images to information entities and extracts attributes such as image size and filename and maps them to repository attributes. Images (photos) may have additional attributes, such as, title or caption, occasion, location, and date and time. However, these are not available, and must be assigned by the user in order to organized the information into a photo album organized in a view by date and time, or a particular location, or a particular occasion, et cetera. This prototype adapter is a simple extension of the file system adapter mentioned above.

7.2.6 Actions

Above it was discussed how the adapter components allow the mapping of information entities from external sources to be mapped into the repository. Also specific examples of adapters implemented in the research prototype were provided. However, the adapters may also map functionality into the system in terms of actions (commands) that can be used in the system. For instance actions that are specifically relevant for at particular information source can be included in an adapter, and may then be added to the system.

The open architecture of the system allows new functionality to be added to the system. This functionality can be made available in the system as commands that can be invoked by the user of the system. An action can be invoked on an information entity or a set of information entities in a view. It may require that specific attributes be assigned to the information entity. The action components and interfaces are located in the *nexus.action* package. Every action added to the system has a name that is used to identify the command, and is used by the user interface of the system.

The research prototype implementation contains a number of actions, which are added to the system via the specific adapters. The adapter providing e-mail mapping to IMAP servers also contains an action *Send* for sending e-mail using an SMTP server. The action can be invoked in principle on any information entity in the repository. However, in order for the command to succeed the information entity must have certain attributes associated. In the case of sending an e-mail attributes indicating e-mail attributes, such as, *To*, *Sender*, and *Subject* must be

present. Invoking the *Send* command on an information entity that has those attributes assigned and provided the *To* attribute contains a valid e-mail address, the command will map the information entity and attributes to an e-mail that can be sent using an SMTP server.

The MP3 adapter described above also includes a couple of simple actions *Play* and *Stop*, which includes the codec to be used to play sound in the MP3 format. The *Stop* action halts an executing *Play* action. The actions are available in variants as they can be invoked either on a single information entity or a set of information entities (provided via a view of information entities). The requirement of the *Play* action is that the information entity contains a stream in the MP3 format; otherwise the action simply has no effect. Similar the *Stop* action has no effect if no *Play* action is currently active.

This action architecture allows any number of actions to be added to the system. General actions may also be added to the system, meaning that they do not have to be coupled to adapter components. However, it is useful to couple actions that are relevant to a specific adapter, as seen in the example above. The actions may be used across all information entities in the entire system. This means that actions have no boundaries to particular information entities even if provided via an adapter. However, the action can only be performed if an information entity has the necessary attributes associated, and thus meets the specific criteria of the action.

7.2.7 Indexing Component

As the repository is meant as a storage mechanism that provides a unified point of access to personal information of many different types it could potentially hold considerable quantities of information entities and attributes. In order to provide sufficient performance when interacting with the system the information in the repository is indexed by an indexing component. This enables efficient retrieval of information from the repository and efficient construction of views of information entities.

Indexing is implemented in the index component that is part of the repository, and is therefore located in the *nexus.repository* package. The indexing component is responsible for indexing of information entities, attributes, and

views. The component therefore holds several indexes of the information stored in the repository.

There is an index of all information entities stored in the repository – the *entity index*. The key used in this index is the unique information entity identifier used generally in the system. The identifier can be used to perform an efficient lookup of specific information entities in the index. Similar there is also an index of all views defined in the system – the *navigation index*. It contains an index of all views with associated name of the view and the attributes that are included in the index.

However, the most important index in terms of efficiency is the index of attributes – *attribute index*. Views are constructed on the basis of attributes associated with the information entities. Thus in order to avoid a time-consuming search of entities the attribute index provides an index of entities and all attributes associated with the entities. This means that a view can be efficiently generated as the attributes included can be directly mapped to a set of information entities by means of the attribute index.

All indexes are constantly kept up-to-date, that is, each time an attribute is added or removed or an information entity changed this is reflected in the relevant indexes. For instance if a new attribute is associated with an information entity the attribute index is updated to reflect this immediately. That is, the particular attribute is now associated with the given information entity. This will also have an effect of the views including this attribute, in the sense that the view might now include the information entity, that is, depending on the actual criteria for the view.

The implementation of the indexes provides sufficient performance for the research prototype, that is, at least for initial experiments to be carried out. The indexing approach involves the classical tradeoff between performance and storage. The cost of the indexing approach is storage and memory, however, with the benefit of efficiency when searching for keywords in the repository and generating views of information entities. This also includes efficiency when changing views so that information entities are included or excluded from the view.

Finally there is also an index of templates available in the repository. Each template has an associated name. For instance an e-mail template is an information entity with associated attributes, such as, To, Cc, Subject, and Body. In

these attribute name/value pairs the value is left blank for the user to fill out. The information containing the template name and the actual template information entity with associated attributes is stored in the template index. When constructing an email a clone of the template information entity with associated attributes is created. Thereby a new information entity is created with attributes that can be filled in by the user. All such templates are available in the index of templates.

7.2.8 Additional Components

The main components of the system have been discussed, except the part that deals with the user interface of the system. Those components are discussed in the next section. In addition to the main components there are some smaller helper components implemented in the system. For instance there is a sorting utility implementing the quicksort algorithm, enabling the sorting of different types of information (numbers, text, dates, et cetera). The component is mainly used to allow an efficient way of sorting views by different attributes.

As mentioned the current research prototype represents a pilot implementation. However, it should be mentioned, that the architecture of the system allows changing these components without affecting the rest of the components in the architecture. Thus the system architecture provides a test bed for further development and experiments.

7.3 User Interface Prototype

The remaining components in the system are related to the user interface of the system. All components and interfaces for the system user interface are located in the *nexus.ui* package.

7.3.1 Model-View-Control

The system architecture is built on the Model-View-Control – MVC – design pattern [GHJV95]. This approach provides an appropriate abstraction of func-

tionality and allows the creation of multiple ways of viewing the same information available in the model.

In this architecture the repository serves as the model. All information that is shown in the display is located in the repository (the model). The view is the actual user interface and display the information available in the repository. This includes the view of different collections, and information entities. Finally the *control component* serves as the mediator handling user input.

7.3.2 Screen Shot

A screen shot of the prototype user interface of the research prototype implementation is shown in Figure 7.2.

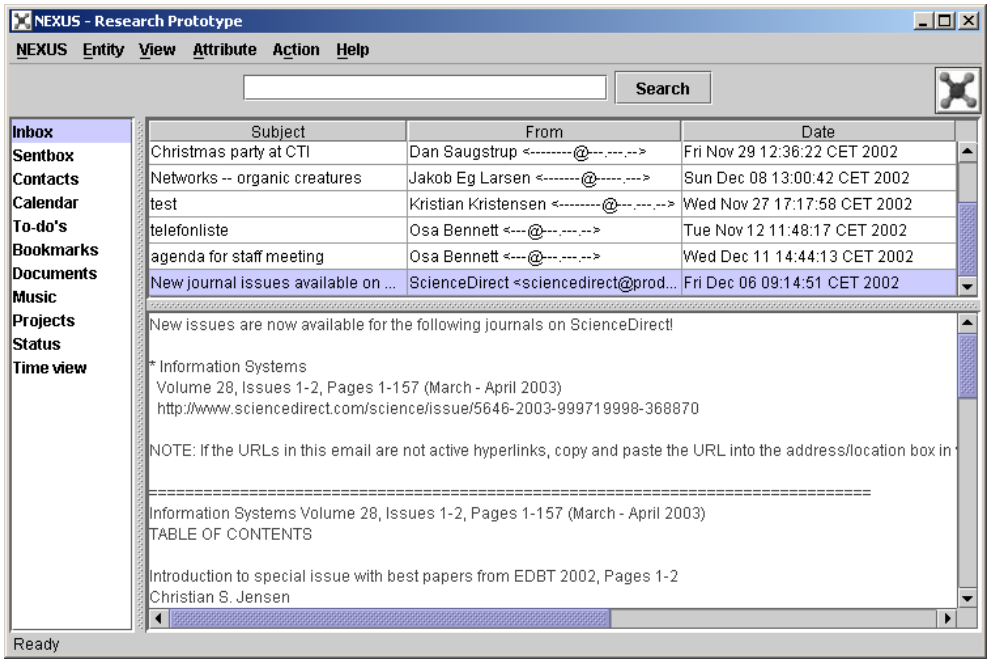

Figure 7.2: Screen shot of NEXUS research prototype user interface

The immediate appearance of the user interface of the prototype has some similarities with standard software applications, such as email applications. However, the underlying model and repository makes it somewhat different. On the top of the display are the menu and toolbar, similar to typical desktop software applications. These are standard graphical user interface elements. The toolbar provide access to frequently used actions, such as creating new information entities based on templates and navigating the user interface.

The remaining display is divided into three main areas. The left area contains a list of views defined in the repository, enabling the user to navigate the different subsets of information entities from the repository. Selecting one of these views affect the remaining two display areas. In the screen shot the currently selected view is the e-mail inbox. The right side of the display contains two parts. The upper part contains the actual view of information entities with associated attributed. In the screen shot above the view contains a list of emails with associated attributes. The lower part contains the content of the information entity. In this case it means the body of an email message.

7.3.3 Views, Information Entities, and Actions

The architecture allows different implementations of views. This means that information can be viewed in different ways. The screen shot shown above shows an example view that provides a tabular way of viewing information entities. On the x-axis are the different attributes in different columns and on the y-axis are the different information entities. Similar a temporal view is implemented for displaying temporal ordered information. This is relevant for calendar and diary information.

As mentioned the architecture allows more kinds of user interfaces for views to be added to the system. For instance it would be obvious to have a spatial view that would enable a spatial organization of information entities in the repository. However, this is not implemented in the research prototype.

The views may allow direct data entry of attributes. In the tabular view mentioned above the user may place the cursor in any attribute field and enter an attribute or alter an existing attribute. Naturally it is required that it is possible to change the actual attributes shown in the view. Some attributes that

are extracted from information entities in from an external information source it is not possible to alter. In that case the user interface does not allow the attribute to be changed. The field in which the attribute is show is simply not editable.

The type of data entry implemented in the research prototype causes the repository to be immediately updated. As soon as the user has typed the attribute into an attribute field it is stored in the repository. There is no need to explicitly save the information entity to the repository. This approach enables the user to quickly enter information into the system. It is not necessary to explicitly name or organize the information entity, as the user can simple perform data entry directly in the view of information entities.

The collections (views) are defined in terms of entities and attributes, which constitutes the structures of the collections (views). Thus the collections are part of the content not part of the user interface as such. This means that the structure is an integrated part of the information in stored in the system rather than being defined by the user interface as such. The user interface views mentioned above are simply different ways of viewing the information in the repository. These user interface elements do not impose any particular structure of the information but different ways of visualizing the information entities. For instance the tabular approach shown above is a general way of viewing information entities provided in the system. It shows a list of information entities and the attributes associated with it. If the information entities have temporal ordering it can be shown in the temporal user interface view. Similar if the information entities belong in a hierarchical information structure the user interface view is visualizing the information entities in a hierarchical tree structure (if such a structure is available for the information entities). As mentioned above also a spatial way of viewing information entities could be implemented in the system, but is not implemented in the research prototype. To summarize the structure of the information is constructed on the basis of the attributes associated with the information entities, which means that structure is part of the information rather than the user interface. The user interface simply provides different ways of visualizing the information entities.

7.3.4 Actions and Navigation

Available actions are invoked from the menu. When an information entity is selected in the view, an action can be applied to the information entity. If a view is selected actions can be applied to the entire set of information entities in that view. This approach means that the research prototype user interface implements a noun-verb-type user interface. The information entity (noun) is selected and the action (verb) is applied to that information entity.

The system provides a unified interface and interaction with information in the repository. It allows the user to combine different types of information in the same views. There is one common set of commands available for navigation and managing the information entities in the system. This means that the user of the prototype only needs to learn one set of commands in order to use and manage different types of information entities in the system. The same commands apply across the entire set of information entities in the underlying repository. This approach aims to facilitate learning and the process of habit formation.

This means it is the same commands that apply for creating, altering, and deleting information entities in the entire system. This is independent of the type of information entities that are present in the views that the user is navigated. Navigating views of information entities is unified as interactions with views in the user interface is the same across the entire system, independently of the actual information entities in the views.

As mentioned previously some attributes acquire information entities stored in external sources are not mutable, and can therefore not be changed or updated by the user. These attributes show up in the system as locked information. Both entities and attributes can be locked, meaning that the system does not enable the user to edit or delete those entities and attributes. This means that also entities and attributes stored in the external repository can be locked. These features also appear in the same way in the user interface of the system.

7.4 Prototype Application

The research prototype implementation of the attribute-oriented model, and the unified user interface is a generic approach. This means that in principle any

kind of information could be stored into the system and it could be applied for different purposes. Naturally the purpose here is to enable personal information management as the unified approach in the model and user interface aims to support the specific issues involved in personal information management.

The types of personal information used with the research prototype for experimental purposes include:

- E-mail
- Calendar items
- To-do items
- Text items
- References (including bookmarks)
- Music

The adapter components enable to bridge existing information sources to the system. The research prototype implement adapters, that are capable of mapping emails, documents and files, music item, et cetera into the repository as discussed in Section 7.2.4 and 7.2.5. As mentioned additional attributes may be assigned, meaning that some attributes are stored in external sources and others are stored in the internal repository.

As mentioned previously the mapping of information entities representing information, such as calendar items, to-do items, text items, bookmarks, and references are not defined by the system. The model provides a generic way of representing such information entities with associated attributes. The information used with the research prototype use simple representations. Bookmarks have associated attributes providing a title, a date, and the URL of the bookmark. Calendar items have associated attributes providing a title, a date, an end date, and description. To-do items have associated attributes proving, a date, a status, urgency, importance, and a description.

The representations mentioned here provide a simple way of representing the different kinds of typical personal information items. However, the user is not

restricted by these representations, as the model allows the values of associated attributes to be empty and also the user can add additional attributes to these information entities if needed.

7.5 Limitations

The research prototype implementation described in this chapter is considered a pilot implementation that serves as a "proof of concept". The implementation includes the core concepts discussed and is based on the attribute-oriented model. The system also aims to provide a unified user interface and interaction with the information in the system.

The implementation is trivial and some functionality is not implemented for instance some user interface views are not implemented. The user interface is simple and provides access to the core elements of the system only. For instance it is possible to create different information structures using the prototype user interface, however the user interface lacks features that could facilitate the process of creating different types of organization. The user interface components for creation of views could be improved. New templates cannot be defined in the user interface of the prototype; instead they are defined outside the system. However, it is possible to add this feature to the user interface of the system.

The prototype is not performance optimized in any way. Performance has only been considered an issue if the implementation yields response times that would prevent experiments to be conducted. However, the research prototype works in a satisfactory way in terms of response times when the information repository has 100s and 1000s of information items with associated attributes. The prototype has not been tested with 10000s and more information items. With the hardware configuration mentioned previously the research prototype has a reasonable performance as it typically has response times below 100–200 milliseconds. The most performance intensive tasks include changing views and sorting views by a different attribute. These tasks have reasonable performance in the system with 1000s of information entities in the repository.

As the research prototype is a pilot implementation and there are a number of limitations the implementation does not represent a system that is ready for

everyday usage. The aim has been to enable initial laboratory experiments to be carried out. These may provide an initial understanding of the application of these principles for personal information management. Additionally it could provide input for further development of the prototype for further research of the area. Also further implementation improvements and optimizations are possible.

The current prototype is also limited as it can be used on a standard stationary or laptop computer system only. Supporting NEXUS on mobile information appliances is an obvious further development of the research prototype. This includes a set of problems to research further. Mobile information appliances, such as, PDA's and mobile phones are typically characterized by limited input and out capabilities, compared to traditional desktop systems and workstations. The output limitations include display space and resolution. Input limitations include limited or no keyboard and maybe a pen as pointing device.

These limitations put a set of constraints on the design of the NEXUS user interface on such mobile information appliances. The display has limited space to present information entities and views and the interaction and data entry are also limited.

The information entities are independent information units. The associated attributes are not part of the units as such but also independent information units. The attributes are coupled to the information entities, but have a representation independently of the information entities as such. The approach facilitates such adapting of the model to mobile information appliances. The limitations imposed by mobile systems means that it could be relevant to leave out some attributes in a mobile situation of use. Associating context dependant attributes, or even context-aware information is receiving attention in ubiquitous computing research [MHP00]. This means that some attributes could by relevant in certain contexts only.

The issue of adapting the user interface of NEXUS to mobile information appliances includes adapting the NEXUS user interface, and the presentation of the individual information entities as such. For instance making available multiple versions of an image or scaling to different image resolutions and possibly both color and black and white versions of the image available.

7.6 Summary

The previous chapter considered interactive personal information management systems in terms of models on three different levels.

First there is a conceptual level, meaning the underlying understanding of the issues involved in personal information management as such. These issues were discussed in Chapter 2 and 3. It is on the basis of this understanding systems are created and existing approaches and systems supporting personal information management were discussed in Chapter 4.

Secondly there is an interaction level, which includes the ways in which users interact with and navigate the information available in a system. Interaction related issues were discussed in Chapter 5, and specifically how to provide unified interface and interaction with information and functionality available in a system. This chapter has discussed a prototype implementation based on these principles.

Finally, the data level includes the actual information representation and structure in a system. Different approaches at the data level were discussed in Chapter 6. The chapter also included a presentation of the attribute-oriented approach on which the research prototype described in this chapter is based.

A simple research prototype implementation has been described. It implements the key elements of the attribute-oriented model and unified approach described in the previous two chapters. The system architecture has been described including the repository for storage of information entities, the adapter framework for enabling access to existing information sources. The description also includes a simple prototype user interface for the system. The user interface allows unified access and use of the entire set of personal information entities in the system. Such information includes messages (emails), contacts, calendar items, to-do items, documents, bookmarks, images, music, and these information items may be obtained from a number of external sources through adapters.

The research prototype discussed in this chapter represents a pilot implementation, which serves a proof-of-concept and a test bed. It includes prototype implementation of the key components onto which a simple personal information management system is implemented. The prototype allows initial experiments to be carried out. As it is a research prototype only it has a set of limitations, which prevents it from being used on a large scale. The prototype does not

implement all elements of the system, thus the current implementation cannot be used as a tool for daily personal information management activities. Implementing a solid and fully functional personal information management system that could potentially be used in a realistic environment is a major undertaking that requires comprehensive systems development effort.

The aim has been to reach a proof of concept level implementation, and to enable some initial experiments to be carried out. This may provide feedback that can be used in further development of the concepts and prototype. Thus the research prototype serves as an experimental platform, which can allow carrying out further research of unified attributed-oriented models and unified user interface for the support of personal information management.

Chapter 8

Experiments

This chapter describes the experimental approach. First experimental dilemmas involved in researching the area of personal information management are discussed including implications for carrying out experiments in the area. This involves a number of trade-offs that must be taken into consideration when carrying out experiments. Also this has implications in terms of applying and generalizing results obtained from the experiments.

The experiment carried out is divided in two parts, which include two different types of experiments. The first experiment involves semi-structured interviews with a number of test subjects to get a qualitative insights and an understanding of their current personal information management practices and work environment.

The second experiment is a pilot usability experiment of the research prototype implementation described in the previous chapter. The experiment is carried out to get initial insights and understanding of the implications of the unified attribute-oriented approach in terms of supporting personal information management.

The overall purpose of the experiments is to get qualitative insights and understanding of the current problems of personal information management and indications of the implications of the unified attribute-oriented approach in terms of supporting personal information management. Each experiment is discussed

further in this chapter. This includes the test subjects used, methodology of the experiments, and experimental designs, which are described and discussed.

8.1 Experimental Dilemmas

Carrying out experiments in the area of personal information management involves a number of experimental problems and dilemmas. These problems involve privacy of information due to the nature of the information and the contexts in which subjects use the information. As a consequence the possibility of carrying out experiments in a realistic setting could be limited. Also carrying out experiments using research prototypes such as the one described in the previous chapter involves a number of problems. In general these issues may impose difficulties and constraints on studies of personal information management.

8.1.1 Ethical and Privacy Issues

First of all there are a set of ethical issues. The information is considered private by subjects and often to be used and seen by an individual only. Subjects way perceive observational studies as an invasion of their privacy. This may impose limitations on the information and management practices that can be observed and may rule out logging of use of personal information management systems. Therefore closely observing actual use of personal information management practices and habits can be difficult. The private nature of information may cause subjects to be reluctant to participate in experiments or withdraw from an experiment if they feel they have to reveal information that is considered private.

Therefore it is important that subjects are carefully informed that the experiments are not concerned with their information as such, but rather a study of information management practices in general. Also that it is not the subjects being tested or asked to reveal information that they consider private, but a study of how information management is carried out by subjects and the tools used to facilitate the processes involved.

8.1.2 Realism and Alternative Tools

The privacy issue may be addressed by using simulated tasks and simulated personal information. However, this is far from an optimal solution as it creates an artificial situation of use. When simulated information is used instead of the personal information belonging to the subject, the test subjects have to put themselves in an artificial situation, which may not entirely reflect their actual personal information management practices. However, when carrying out experiments involving tools or research prototypes that implement an alternative approach to the support of personal information management simulated tasks and information may be a necessity.

An essential problem in personal information management is the role of memory and recollection of personal information. Naturally subjects have no recollection of simulated personal information in a system. The individual has not handled the information before, and thus an essential aspect of personal information management is lost when using simulated information. Nevertheless experiments can provide some indications of use and may reveal problems in personal information management tools.

In terms of real use or simulated use at least three types of personal information management experiments can be carried out:

1. An experiment involving observations of real use: Subjects would perform real task using the tools and information used in real situations.

2. An experiment testing a new personal information management tool in a realistic setting: Subjects performing real task using real information, but using a new personal information management tool.

3. An experiment testing a new personal information management tool in a laboratory setting: Subjects performing simulated tasks using simulated information and using a new personal information management tool.

Each of these experiments mentioned above would provide different types of results. The first could provide an understanding of the current information management practices and potential problems in the tools or practices impeding the processes of performing personal information management. The second

could provide an understanding of the implications of an alternative approach to supporting personal information management by means of an alternative tool. However, it does require the alternative personal information management tool to support the information used and tasks carried out by the subjects. The third kind of experiment could also provide an understanding of the implications of an alternative approach to the support of personal information management by means of an alternative tool. However, this includes the limitations imposed by introducing simulating tasks and information in the experiment. It also depends on the type of information and tasks that can be carried out using such a prototype tool. This again depends on the scope of the prototype system.

8.1.3 Experiments Involving Prototype Systems

In the second and third type of experiment mentioned above the scope of experiments that can be carried out depends on the scope of the prototype tool. An extensive prototype may allow extensive experiments involving everyday tasks and information. Whereas a limited prototype may hinder the experiment if the prototype does not support the kinds of activities carried out by the subject. Thus a simple prototype may only provide initial indications of the implications of an alternative approach.

Experiments in a realistic setting require the subjects to perform real tasks and carry out real use. This means that a prototype system for personal information management needs to be extended to a level where it can support everyday tasks and information management. This means that extensive development effort is needed to develop the research prototype to a level where it supports the information used and tasks carried out by test subjects.

As mentioned above the role of memory is essential in personal information management. This means that in order to test out the implications of an alternative approach in terms of supporting memory (recollection and retrieval) ideally longitudinal studies are needed. Otherwise it can be difficult to reach conclusions about the support of memory, as naturally some time has to pass to observe the retrieval and reminding support. That is, the subjects would need to manage the information over some period of time by use of the alternative approach (e.g. a research prototype) and then at some point later in time use

it to retrieve the information. Thus the tool becoming a part of the everyday personal information management related activities. Studies that require use of an alternative approach over longer periods of time require the willingness of subjects to invest the time needed in order to use such an alternative approach (e.g. a research prototype system).

Carrying out limited laboratory experiments that involve a different approach is then a possible alternative. Such studies including simulated personal information management tasks can provide some valuable indications of the use and the value of the approach to personal information management. However, in terms of personal information management issues such as archiving information and reminding it can be very difficult to reach results. Studies over longer periods of time are needed to get useful results on the support of archiving, retrieval, and reminding.

Experimental dilemmas	Simulated situations of use, tasks, and information	Realistic setting, context, real use, and real information
Prototype system	Provides indications of use user interface problems, and indications of the model supporting personal information management.	Real use of prototype system. This may provide useful results about the approach. Requires extensive development of prototype.
Extensive System	Enables more tasks and situations of use to be tested. Simulation is a limitations to an experiment.	Enable observations of current practices and may provide stronger results. Does not test out an alternative approach.

Table 8.1: Experimental dilemmas of using 1) real versus simulated information and tasks, and 2) using real versus prototype systems for experiments

The experimental dilemma is related to the extent of development needed on a research prototype in order to allow experiments to be carried out in a realistic setting based on real use and management of subjects own personal information. An alternative is a simple proof-of-concept research prototype, which may allow

experiments involving simple simulated tasks and information to be carried out. Such experiments may provide initial indications of use, user interface problems, and the underlying model as a foundation for the support of personal information management (see Table 8.1).

8.1.4 Incremental versus Novel Approach

Research of personal information management can be approached with different types of experiments. Generally introducing a different approach and testing out a research prototype implementing the approach can demand considerable resources. It requires extensive development effort in order to enable realistic experiments to be carried out. However, an experimental dilemma is that a different type of approach forces subjects to take the effort to learn a different tool and different approach to personal information management. This may be difficult for subject as they may already have well-established information management practices using a set of existing tools to facilitate the processes involved. Their practices may also to some extent be formed by the specific constraints of existing tools used [FFG96a].

To overcome this issue an alternative approach is to use an existing system with some incremental changes instead of a novel research prototype system. The incremental changes to an existing system can be used to test out specific alternative features of an approach to personal information management.

The advantage to the *incremental approach* is that the tools used in experiments are then familiar to the subjects [Ras94]. The subject will only need to learn the incremental changes to the existing systems in order to use these features. Also it may require less development effort to change an existing tool compared to developing an entire new system. On the other hand the experiment can only be carried out within the boundaries of the existing tool. This means that the changes can only be done based on the model of the existing tool and within the constraints in the existing system. Also some systems may prevent such development or customization thereby preventing such incremental changes to be implemented for research purposes.

Using the *novel approach* means that a new system with a different approach to personal information management is used in the experiments. This means

that subjects need to use a new tool that is unknown and perhaps quite different from the tools they are currently using. They will need to learn how to use and become familiar with a new tool, which means that subjects need to invest time and resources doing this. Also this potentially demand longer-term use in experiments compared to the incremental approach. The novel approach may also require extensive development effort to create a working prototype, which may be used for carrying out experiments. However, the advantage is that there are no constraints by an existing tool, which is the case in the incremental approach. This leaves room for more flexibility, in terms of a prototype testing a new approach. The dilemmas of the two approaches are summarized in Table 8.2.

	Incremental approach (extension to existing system)	Novel approach (new system and new approach)
Tool used	Familar tool	New tool
Differences from existing system	Incremental changes	May be very different
Development effort	May require less development effort to extend existing system	Extensive effort to develop a working system
Implementation constraints	Within boundaries of the existing system	Can test out different novel approaches without constraints of an existing tool

Table 8.2: Tools and development of tool for experiments

Conducting experiments using the *incremental approach* has the advantage that a familiar existing tool is used. As the changes to existing tools are incremental, the learning of the feature might not require much effort by the subject. This also means that it can be possible to carry out the experiments in a "real" workspace and environment. Potentially the subjects can be carrying out real activities in a familiar environment and apply the incremental features in their activities. Thus studies may even contain studies of real use, in real situations,

and using real information. Also studies over longer periods of time are possible as subjects are using a familiar everyday tool. However, a problem compared to the novel approach is that the findings in the incremental approach may be results on the particular feature that has been modified. This may be a finding on a fairly detailed level, as the change is a modification of an existing tool. In terms of testing an alternative approach to personal information management support the incremental approach only test modifications of an existing tool and does not provide information whether a different approach might be advantageous over the existing approach. Perhaps the approach can reveal some problems in the existing approach and incremental features that provide hints whether an alternative approach is worth pursuing. On the other hand it may possible to draw fairly solid conclusions about the particular feature being tested in the particular context.

Experiments using the *novel approach* have a set of difficulties compared to the incremental approach. As the approach being used in the experiment is novel, it means that users have to use a potentially widely different tool compared to the tools they are familiar with. They have to learn the novel approach to an extent so that they can apply the system to support their personal information management activities. As the tool is novel it may not be possible to apply it in a real situation. Thus experiments may have to be carried out as laboratory experiments, which means that subjects are taken out of their real work environment. Thereby they are put in an artificial situation of use. Additionally laboratory experiments means "simulated" tasks with simulated information, which involves the problems and issues discussed in the previous section. Test subjects have to identify themselves with the simulated task, situation, and information. As mentioned previously studies of longer periods of time are difficult, as the experiments are carried out in a laboratory setting. Alternatively a functional prototype is required for longer-term experiments and real use situations. However, this also means that a fairly functional prototype tool must be developed, in order to conduct such experiments. As mentioned previously such development efforts can be substantial. A functional prototype is needed if subjects will be using the tool to test the approach to a degree that enables solid conclusions about the approach. However, a limited prototype implementing a novel approach may be able to produce initial indications of the possibilities of the novel tool or approach to personal information management

support. Thus it enables some results to be reached and limited conclusions to be made unless extensive real use is possible using a functional prototype. The dilemmas of using these two approaches are summarized in Table 8.3.

	Incremental approach	Novel approach
Experimental tasks	Can be carried out in "real" workspace and environment	May have to use laboratory experiment
Realism	Real use possible with use of extensions to existing tool	"Simulated" tasks and information (in laboratory experiment)
Time period	Using everyday familiar tools, thus longer time periods possible	Laboratory setting may prevent experiments over longer time periods
Learning	Incremental to existing tool, needs to learn incremental features only	Have to learn new tool that may be very different from existing tool
Results and findings	Results on the incremental feature only, not about a different approach	Indications of alternative approach
Possible conclusions	Conclusions about a particular incremental feature (in the context of an existing system)	Potentially limited results and conclusions unless extensive real use is possible

Table 8.3: Experiments using incremental versus novel approach

8.1.5 PIM Research Methodologies

Some of the experimental problems discussed here are specific to studies of personal information management. For instance the privacy and ethical issues mentioned are considerable in personal information management. However, the problems related to the experiments using a research prototype involve general

experimental problem in human computer interaction research.

Research of personal information management has revealed a set of common problems in information organization and retrieval, as discussed in Chapter 2 and 3. Over the last decades novel solutions and radical ideas compared to existing systems have been proposed in order to address these problems [WTN00], examples of such novel solutions include Data mountain [RCL+98], Lifestreams [FFG96b], PAD Zoomable interfaces [PF93], see also Chapter 4. However, most often only limited experiments, evaluation, and user testing of these novel approaches have been carried out. Of course these approaches have to address the experimental problems mentioned above. Moreover no attention has been on comparative experiments of these approaches [WTN00]. Each approach may claim to have addressed important issues in personal information management. However, personal information management activities are highly complex and thus an approach may only provide valuable support of a subset of the activities and aspects of personal information management. Other aspects and activities may have poor support or none at all.

Experiments are needed in order to clarify to what extent the different approaches address the personal information management issues identified. More research and experiments are needed in order to clarify this. That is, to assess to which degree an approach provides support for the activities and different aspects involved in personal information management. Comparative studies would need a set of methodologies in order to be approached in a consistent way. However, the field has to further develop the understanding of personal information management and the problems involved. This is needed if common tasks should be established and used as the basis for conducting consistent experiments, and conducting meaningful comparisons. Such methodologies have yet to be developed in the research area [WTN00].

The research prototype described in the previous chapter represents a pilot implementation of the key components of the unified attribute-oriented approach to personal information management. It allows initial experiments to be carried out, as the research prototype has a set of limitations that prevents it from being used on a large scale as a tool for supporting daily personal information management activities. In terms of the experimental dilemmas initial laboratory experiments may be carried out. This means initial experiments with simulated

information and tasks. However, the experiment has been divided into two parts – a semi-structured interview and a pilot usability experiment.

8.2 Experiment 1 – PIM Practices

The first experiment aims to obtain a qualitative insights and an understanding to the way in which subjects perform personal information management in their existing environment. The purpose of the experiment is to obtain an understanding of personal information management practices and issues and the problems involved in personal information management activities. The aim is also to identify problems in current interactive systems used by the subjects to support their personal information management practices.

Additionally the purpose of the experiment is to study if the problems identified in prior studies exist with the use of present tools for the support of personal information management. This means to get a confirmation of previous results, that is, to verify whether they also apply in present information environments. Information environments have changed drastically over the last decade especially due to the availability of vast information via networks. Thus additionally the first experiment serves the purpose of getting a further understanding of personal information management practices and requirements for the support of personal information management in interactive systems in present information environments.

Prior research of personal information management has typically focused on a single application or information type. For instance study of email management [Mac88] [WS96] or bookmark management [ABC98]. Contrary this experiment aims to take a unified view on personal information management instead. Rather than studying the use of a specific type of information or application, such as email use or bookmark management, the first experiment aims to address the information environment used by subjects as part of their daily information management activities.

8.2.1 Subjects

Eight people were interviewed about their information management habits and practices. The subjects were chosen to represent different people who do most of their work at desks, and having a considerable use of information in their daily activities. Additionally having a fairly extensive use of computer-based tools in their daily activities in order to support personal information management related activities. This means users of interactive systems in which they manage personal information.

The subjects interviewed were all male age 25–55. They included two consultants, two academic researchers, a systems administrator, two master level engineering students, and a self-employed businessman. All subjects except one did the main part of their work at desks and were dependent of different types of interactive systems, tools, and information in order to carry out their daily work activities. The self-employed businessman was an exception as he did approximately half his work at the desk and the remaining work out of the office.

8.2.2 Methodology

The experiment was carried out as semi-structured interviews, see Appendix A. The interviews took place in subject's own office environment. The purpose was to perform the interview in their familiar information environment at their workplace. This enabled subjects to point out items and tools used as part of everyday activities. The information environment included all information in the office including both paper-based and electronic-based information. Thus it also included the desk, shelves, bulletin board, whiteboard, et cetera.

The interviews took approximately one hour. Some subjects were very interested in telling about their work and practices, which made the interview take longer (in one case up to two hours). Others subjects were a bit reluctant to reveal their personal information and practices.

The experiment aimed to research the information environment and the personal information management procedures and habits. The interviews were carried out as semi-structured interviews where the interviewer tried to get the subjects to tell as much as possible about their personal information environ-

ment and personal information management practices. Initially the subjects were asked to describe their work environment (the desk) and the tools used to manage their information. This included the physical desk as well as how they organize information by means of different interactive systems and tools. During the initial part of the interview the subjects were encouraged to tell as much as possible. For instance the interviewer asked them to think about things they would do in a typical day, or what they did earlier the same day. Initially the subjects were encouraged to give a tour of the office and tools used on a daily basis.

Generally the interviewer aimed to ask open-ended questions allowing the subjects to provide as much information as possible. The initiative was given to the subject when possible in order to encourage them to talk freely about their information management practices.. In some situations the interviewer would ask questions in order to get clarification. In other situations the interviewer supplemented by follow-up questions encouraging them to continue and provide further details. Such questions included "What is this?", "How/when do you do this", and "Why is this information here?".

During the interview the subjects would point at things in the office and information environment. Having the interviews in their usual work environment also allowed to pose questions such as "What do you keep in this stack over here", "When do you use these document?", "What's the purpose of the notes here", and "How often do you use the stuff on these shelves?".

After the initial part subjects were asked more specific questions about email use, bookmarks, calendar use and management practices. Additionally they were encouraged to tell about problems and difficulties the tools or features they thought they were missing. Subjects were not asked any structured questions as such during this part of the interview.

In the end of the interview subjects were asked a set of questions in order to encourage them to talk further about their practices and possible problems involved in their information management practices. These questions included:

- How often do you have problems finding the information that you are looking for?
- What do you think the reason is if you cannot find the information?
- What are the biggest problems organizing your information and why?
- What do you think is the biggest issues in your current approach to organize your information? Is there anything you would like to do in a different way?
- Do you have suggestions to how the problems could be solved?
- Do you think you need additional tools in order to address the problem?

Finally the subjects were asked to give an estimate of how structured they considered their information organization. The estimate was provided on a scale from one to five, see Table 8.4.

One	Very good
Two	Good
Three	Neutral
Four	Poor
Five	Very poor

Table 8.4: Scale for information organization practices as indicated by subjects

The interviewer wrote down the answers from the subjects, including spontaneous utterances and actions they would make during the interview. For instance this occurred when demonstrating information management using interactive tools, or when pointing to information items in the office environment. Such utterances were not answers to questions or the interview as such. For instance during demonstration of their information environment they would find items

and say things like "what is this doing here?" and "I need to clean up this". One subject actually briefly started to clean up his bulletin board during the interview when asked how the bulletin board was used. That is, attention was brought to the bulletin board and he realized it had not been cleaned up for a while.

Additionally the interviewer observed the information environment, and notes were taken on the appearance of various information items. This included a brief description of the office and its appearance. That included notes on piles and books on the desk, shelves, bulletin boards, white boards, and other places in which information items were kept. Additionally notes were taken on the tools used for handling and managing information in the office environment. This included notes on computer-based tools, such as, PC, PDA, mobile phone, et cetera, and paper-based tools, such as, calendar, schedules, notebooks, et cetera.

8.2.3 Limitations

A possible limitation to the experiment was that the subjects did not perform actual tasks while being interviewed. The experiment did not include observation of actual tasks being performed as such. Rather the subjects were only demonstrating how they would use and manage their information. Potentially there is a gap between how they talk about their activities and how they actually perform their activities on a daily basis.

On the other hand some subjects were willing to show their actual information organization schemes and explain why they have chosen to organized the information in they way they did.

8.3 Experiment 2 – Pilot Usability Experiment

The interview in the first experiment provided a brief understanding of the current information management practices of the subjects. That is, it provided an introduction to the type of information that was dealt with on a daily basis and the strategies and practices used in order to manage the information. Additionally it gave insights to some of the problems and issues involved in carrying their

activities and personal information management.

The interviews were followed-up by the second experiment, that is, a pilot usability experiment evaluating the research prototype described in the previous chapter. This experiment is carried out in order to get initial understanding and insights into the unified attributed-oriented approach in terms of its support of personal information management. That is, to obtain some indications to what extent it addresses the issues identified in the previous experiment and in previous studies of personal information management.

Thus the motivation of the study and experiments is to apply an attribute-oriented data model in the domain of personal information management, in order to study how interactive systems support the characteristics of personal information. The experiment is therefore considered a pilot usability evaluation experiment. However, the purpose is not to carry out a usability evaluation of the research prototype identifying specific user interface problems. Rather the purpose is to research into the underlying model and unified interaction. The intention is to get initial indications of the usefulness of the model for the support of personal information management. This means the usability experiments aims at obtaining insights to the implications of the model and the unified approach rather than evaluating the user interface of the research prototype as such. Additionally the aim is to establish suggestions on how to carry out further research of personal information management and provide suggestions for further development of the approach.

8.3.1 Subjects

The subjects used in this experiment were the same as the ones used in the interviews in the previous experiment. The previous experiment was carried out in order to get an understanding of the tasks, activities and practices used by the subjects. This provided a background on which to contrast the results from this experiment. That is, it is possible to relate their reported personal information management activities to how they react to an alternative approach to personal information management and the research prototype.

As mentioned previously the subjects were chosen to represent different types of people who do most of their work at desks, and having a considerable use of

information in their daily activities. Also that they already had a fairly extensive use of computer-based tools in order to carry out personal information management related activities. Thus the subjects were considered to represent potential target users of this type of approach to the support of personal information management.

Using the same subjects in the second experiment could potentially be considered a methodological problem. It could be argued that the subjects participating are biased by the interview and the emphasis on problems in current tools during the interview. However, as the aim of the explorative experiment is to get initial insights only this problem was considered to be limited. Also the purpose was to contrast their reactions to the model and research prototype against the information they had provided about their personal information management practices.

Using eight people for the experiment may be considered a small sample size. However, studies have shown that five test subjects will typically reveal about 80% of the problems in a design being evaluated [Nie93b]. Secondly the limited functionality of the research prototype prevented extensive experiments to be carried out and involving real situations of use. Thus eight subjects are considered sufficient for carrying out a pilot experiment.

8.3.2 Methodology

The methodology chosen for this experiment is based on a usability experiments carried out as simple thinking-aloud experiments [Nie93b] [Jør90]. The subjects were provided with a few simple tasks to carry out using the research prototype, see Appendix C. In the process of carrying out the tasks, the subjects were asked to think-aloud. They were asked to utter any thoughts coming to their mind while doing the tasks, including any uncertainties, doubts, or surprises. That is, if the research prototype responded in ways different from what they expected.

As mentioned above, due to the limitations of the research prototype the real type of tasks that can be carried out using the prototype are limited. Instead subjects were carrying out simple tasks using the research prototype system. They were asked to imagine that they were using the research prototype to

manage their personal information. This would include storing emails, address book, bookmarks, to-do items, calendar appointments, and documents into the same system – the research prototype system.

They were provided with a brief introduction to the system that included a brief explanation of the key elements of the model, see Appendix B. That the information is represented by items having associated attributes. The information is organized in the system into different views and that everything is accessed using the same user interface of the research prototype. Also that commands can be issued on the information items using the "Action" menu.

As the former experiments were carried out in the information environment (office) of the subjects also this experiment was carried out in that environment. The advantage was that subjects did not have to go into a laboratory for the experiment. Instead an experimental machine that were capable of running the research prototype was brought into their familiar environment. However, using the research prototype, simulated information, and tasks can definitely be characterized as a laboratory experiment even though carried out in their familiar environment.

The research prototype system had a default repository with typical personal information, such as, emails, to-do items, calendar items, documents, music tracks, et cetera. The repository used in the experiments held approximately 500 such information entities. This was considered a realistic quantity of information that could enable the subjects to put themselves in a situation where they would be managing this information. Naturally the actual quantities of personal information kept by subjects are highly individual and varying. As discussed in the beginning of this chapter an experimental dilemma is that this information kept in the repository was not personal information belonging to the subjects. Moreover the subjects did not organize the information into the system. However, the purpose of the experiment was not to study how the subjects would manage their own personal information in the system, but rather to get some initial indications of the use of the system approach in terms of supporting personal information management.

The introduction and the simple tasks should give the subjects an understanding of the model and approach to personal information management.

- Check the email inbox
- Associate an email to a particular project
- Create email
- Send email
- Navigate the bookmarks
- Find and access a bookmark
- Navigate the music in the system
- Start a music track

See also Appendix C.

An issue to deal with in carrying out such experiments is to choose the right granularity of test tasks and activities for the subjects to carry out. It could be argued that the test tasks should correspond to tasks that the subjects would normally carry out during everyday situations of use where personal information management is carried out. Creating such tasks and activities that corresponds to the tasks carried out by the subjects on a daily basis is difficult, as they have highly individual practices. Though the subjects were in an artificial situation of use, they were able to relate to the problems of managing and retrieving the kind of information that was kept in the system. The issue was also addressed by asking additional questions during a short debriefing session at the end of the experiment.

As mentioned during the thinking-aloud part of the experiment the test subjects were performing a set of simple test task using the research prototype implementation. The subjects were asked to think-aloud and encouraged to utter any problems encountered performing the tasks. It was observed how they performed the activities and tasks, and the problems they would encounter doing this. The observer took notes of the observations, including utterances by the subjects

and problems encountered when performing the simple tasks. The purpose of the experiment was to get initial feedback and insights into the research prototype and the underlying model for personal information management. Therefore the attention was on problems related to the model and approach rather than specific problems in the simple user interface implemented in the research prototype. The method should therefore also be considered a step in performing an iterative design of a more extensive system that could potentially be used in real situations of use.

8.3.3 Debriefing

The experiment ended with a short debriefing session. This was meant to be a short follow-up discussing of the approach and the problems encountered by the subjects. The subjects were asked to provide any comments they might have regarding the research prototype, the approach, the information and tasks provided, or the experiment in general. See Appendix D.

Although this was initially meant as a short session for clarification, this part of the experiment turned out to be a quite fruitful some subjects provided good feedback during this session and it was possible to discuss the issues in retrospect. Some subjects were very interested in the approach and therefore provided interesting and fruitful feedback during the debriefing session. This meant that they would also comment on things beyond the specific tasks and even relate the approach to their everyday tasks and strategies for performing personal information management.

8.4 Summary

This chapter has describes the experimental approach taken. The experiments carried out consisted of two different experiments. This included semi-structured interviews to get insights to the personal information management practices carried out by the subjects. Additionally a pilot usability experiment involving the research prototype was carried out in order to research into the implications of the unified attribute-oriented approach for personal information management. The subjects and methodology of the experiments was described in this chapter.

The chapter also emphasized the experimental dilemmas involved in research of personal information management.

Chapter 9

Results and Discussion

This chapter presents the results from the experiments described in the previous chapter. The findings on personal information management practices gained from the semi-structured interviews are presented and discussed. Similar the results from the pilot usability experiment carried out using the research prototype are presented and discussed. Finally some suggestions for establishing further research in the area of personal information management are discussed, based on the experiences from this study.

9.1 Results: Semi-Structured Interviews

The semi-structured interviews provided good insights to the current information practices used by the subjects. In the past two decades studies similar to this study have been carried out. However, over the last decade new tools for support of communication and information retrieval has become widespread. Most remarkable is the use of email for communication and World Wide Web as an increasingly important source of information. Therefore a secondary purpose of the semi-structured interview was to get some indications of current information management practices and do cautious comparisons of these findings to prior studies.

As discussed in the following sections the interviews generally confirmed findings of earlier studies. Thus it indicates that problems identified in earlier studies of personal information management still apply with the personal information management support provided by current systems. In general there are some indications that new systems and approaches used currently has not changed the fundamental problems of personal information management that has been reported in research over the last couple of decades.

9.1.1 Individual Style

The first observation that is apparent from the interviews with subjects is that personal information management practices are widely varying and individual styles are noticeable. Some subjects had strong systematic approach in their information management practices, whereas others did not. One subject even described his personal approach as being chaotic. The individual differences among subjects were observed in two ways. One was the indication of organizational practices used by the subjects. The other was based on observation of the appearance of the physical office as well as the strategies for organization of information supported by various interactive systems.

The observations of the office environment gave good insights to individual style used for personal information management. Some offices where characterized by neatly organized information in ring binders organized in shelves and only few papers, books, and other information on the desks. Other offices where characterized by many loosely organized piles of paper on the desk, books, papers, and other information scattered in the office environment. Some subjects had very large piles of paper on the desk, and shelves that appeared very messy. Based on these observations of office environments the subjects could definitely be divided into two different groups similar to those provided by Malone [Mal83]: the neat office and the messy office. This is a rough way of categorizing the office environments and a few subjects could not easily be categorized into one or the other.

Additionally subjects were asked to judge how structured they considered their approach to personal information management. The level of structure indicated by subjects varied between 1 (best) and 5 (worst) among the subjects,

see Table 9.1.

1 very good	2 good	3 neutral	4 poor	5 very poor
1	3	2	1	1

Table 9.1: Indications of how structured subjects considered their personal information management practices

The purpose of this was not to get quantitative measure, but rather to get an understanding of subject's own understanding of their personal information management practices. A couple of the subjects that indicated organization level 2 said that they meant that they were fairly structured in managing their information. However, the reason their reply was 2 was they felt their practices could be improved a little in order to keep and organize their information better.

One subject hesitated when providing an answer to the question. His office was quite messy and he also ended up indicated being unstructured (4). However, he added that even though he knew the office was messy and that he was quite unstructured he somehow always knew where to find his information. The point is that the differences mentioned here do not indicate the efficiency of the individual personal information management strategy. The subjects in the study had different information needs and different tasks were carried out. Thus any such comparisons would be inappropriate.

Generally the impression of the office in terms of being neat and messy corresponded to the indication provided by the subjects. That is, subjects indicating structured also had a neat office, and subjects indicating unstructured appeared to have a messy office. Again there were a few subjects that could be characterized as being somewhere in between.

The study clearly indicated highly individual style and personal information management practices. This included different levels of use, and different strategies in handling the information. Another way to characterize the subjects was their preference for filing versus piling strategies. That is, the ones with neat offices could be characterized as using mainly a filing strategy, whereas the

ones with messy offices could be characterized as using mainly a piling strategy. However, the semi-structured interview revealed that they typically used a combination of the two approaches.

Finally, a couple of external factors were observed as well. It was observed that the individual style was in some cases influenced by company rules and politics. In a few cases company policy required the employees to leave a clean desk at the end of the day. Such policies could possibly impose a person to work in a more structured fashion than would otherwise be the case. The subjects that could be characterized as having a messy office were not subject to such rules or politics. The point is that in some cases it was observed that external factors had some implications on the personal information management practices.

9.1.2 Tools Used

Besides the apparent differences in level of structure of the information practices there were other individual differences as well.

The subjects that were selected for the experiment did most of their work at desks and had a considerable use of information in their daily activities. They used different computer-based and paper-based tools in order to keep and manage their information. All had a fairly extensive use of computer-based tools supporting personal information management as part of their daily activities.

The interviews revealed that subjects used different interactive tools to management their personal information in different ways. The use of information and various tools are summarized in Table 9.2.

Generally subjects indicated to use only a fraction of the functionality available in the interaction information tools. For instance in the use of handheld information appliances such as PDA's and mobile phones the subjects indicated that they only used a fairly limited functionality and information in those appliances.

All subjects had a mobile phone, which were generally used primarily for calls and messaging (SMS). Only two of the subjects used more functionality in the phones. One used it to hold reminders, saying that it was the only device that he would always carry, why it was suitable to keep reminders in the mobile

	1	2	3	4	5	6	7	8
Electronic information								
Emails	x	x	x	x	x	x	x	x
Contacts/addresses	x	x	x	x	x	x	x	x
To-do items			x	x				x
Documents	x	x	x	x	x	x	x	x
Bookmarks				x	x	x		x
Calendar items	x		x	x				
Music items	x		x	x				x
Images	x			x	x			x
Information appliances								
PC	x	x	x	x	x	x	x	
Laptop	x		x	x				x
Mobile phone	x	x	x	x	x	x	x	x
PDA	–		x	–				x
Paper-based systems								
Notebook (diary)	x	x				x	x	
Calendar or organizer	x	x			x	x	x	x
Bulletin board		x		x		x		
Shelves (paper archive)		x	x	x		x	x	

Table 9.2: Indications of information and tools used by subjects
(x) indicates use, (–) indicates partly-used (seldom), and blank indicates no use

phone. Even though some subjects had mobile phones with personal information management features these features were hardly used. One subject commented that he had tried using the calendar functionality but had given up, as it was too difficulty to use and not worth the effort. Four subjects owned a PDA but only two of them used them on a regular basis. The information kept in the PDA was appointments and other calendar information. Thus the subjects in the experiment could be characterized as having limited use of PDA's.

All subjects except one used desktop computer systems on a daily basis. The one subject used a laptop computer instead. All subjects used email, and indicated that they needed email as part of their daily activities. Also all subjects carried out some level of document management in their desktop environment. Only a few subjects kept specific to-do items in their electronic system in order to support reminding. However, the reminding function was then supported in different ways, as discussed in the next sections. Subjects' management practices and use of the different information types are discussed further in the next sections.

As mentioned the subjects in the study were extensive users of desktop systems and electronic information. However, also paper-based systems were used to some extend. Especially dissimilarities in the way calendar information was handled were observed. On the one end of the scale subjects used paper-based calendars only. One subject mentioned that the calendar was his single most important tool for his daily activities and that it would not be impossible to carry out the day without the calendar. All his appointments and other valuable information were written into the calendar. Three subjects reported using both paper-based and electronic-based calendar systems. For instance the electronic calendar would be used for all business appointments and the paper-based calendar for both business and private appointments. Only two subjects used electronic based calendars only. Additionally half the subjects used some kind of paper notebook or diary to keep information.

All subjects used paper-based information in some way or another. In each office that was observed in the experiment some kind of paper-based system was observed. The majority of subjects had some kind of paper-based archiving system where they kept the paper-based information at some level of structure. Other subjects did not have a paper archive as such. At least their use of paper was limited so they only kept a limited amount of paper-based information for longer periods of time. One subject kept a large paper-archive with information that was several years old. However, he reported that use of information from the archive was extremely rare. Finally, in the office environment shelves and bulletin boards were also used to keep information.

The choices of personal information management tools also illustrate the highly individual styles and practices. Additionally the different types of infor-

mation used by the subjects illustrate the differences in information needs and requirements.

9.1.3 Information Types

The observations reveal a diversity of information and information types being used in order to carry out daily activities. A list of some of these information types can be seen in Table 9.2 above. The information was used to support a variety of activities that were carried out on a daily basis.

All subjects were users of email as an integrated part of communication in their everyday life. As part of their e-mail system subject had lists of contacts. It was interesting to observe that subjects had multiple such lists of contacts and addresses kept for different purposes. That is, one containing e-mail addresses (typically associated with the e-mail system), and another address book (paper-based or electronic) to keep physical address information and phone numbers. Some subjects reported to have several such address lists kept for different purposes.

Additionally all subjects kept some level of document archive and therefore needed some kind of document management system. Different strategies and tools were used to keep such documents, again underscoring the high level of individuality in personal information management. Similarly half the subjects used computer-based tools to keep music and image archives. Again different strategies and tools were chosen in order to manage the information. Some used the native file system to manage and organize music and images, whereas others used specialized tailor-made tools to manage music and images files.

It was also found that some subjects related information across activities. For instance a couple of subjects had an extensive use of the e-mail system, and used it for additional purposes beyond e-mail. It was observed that some subjects also used e-mail management system to keep other types of information. Similar to what has been reported in other studies [Mac88] it was found that activities such as, task management, reminding, and document management was carried out using e-mail systems. For instance subjects used the archiving facility in the e-mail system to store documents rather than (or supplementing) storing documents in the file system. Generally there were indications that e-mail tools

played a central role in the coordination and management of several different types of personal information management. Two subjects reported that they preferred one tool (the e-mail system) that provided access to and management of their information. Similar to previous studies [Mac88] [BDHS03] this study also indicated that e-mail systems have become an important tool fulfilling this requirement. Overall the study confirms that a diversity of information is used to support a variety of activities that is being carried out on a daily basis.

9.1.4 Information Units and Structure

The information units that were used by the subjects included the different information types mentioned above. For structuring of information different units were identified. In the process of structuring the information in terms of collecting information in groups, two different structural approaches were identified. As mentioned previously some subjects used unstructured piles of documents as a way of organizing the information. Others used more formally filing of information. Thus piling and filing strategies were also identified in this study.

Moreover the information used by the subjects where organized in different ways. The structuring of information included:

- Unstructured information, which included information that was not put into any formal information structure as such. Examples included information left in physical or virtual inbox, information left on the desktop, or information left in random piles. For some piles subjects were not capable of providing a good description of the content of the pile. The piles were reported to hold miscellaneous information.

- Loosely or semi-structured information, which included information that was organized in semi-structured piles. That is, even though the documents in piles were unordered each pile could be clearly identified and provided a heading by the subject. Some examples reported by subjects included piles of action items, information gathered for a particular project, official documents, reading material.

- Formally structured information, which included information that was archived in some formal way. For instance one subject had an extensive hierarchy of emails stored in his email system. The subject reported to store almost all e-mail that he received in this hierarchical information structure. Also a few subjects kept paper-based information organized in formally structured archives.

In addition to the different levels of structuring of information reported above, the studies identified different classes of information use as well. The information use reported by subjects could be classified into three different sets, similar to the way it has been reported in prior studies [Col82] [Mal83] [BN95]. Information was handled and structured in different ways which can be described as three levels of interaction with the information:

- Ephemeral information – was found as information that would be used for a short period of time only. Thus it would not be stored in any formal way but typically deleted shortly. Examples of such information included e-mails received that were simply read and immediately deleted. Another example is notes written on a piece of paper for instance during a phone conversation and discarded quickly.

- Working information – was found as information that were used presently as part of ongoing activities. This was observed as the immediate active documents on the computer screen, and documents and piles kept in the immediate information environment.

- Archived information – was found as information filed for possible later usage. All subjects had paper-based and/or electronic archives.

9.1.5 Storage, Retrieval and Classification

The experiment also provided indications of storage and retrieval issues. As mentioned previously subjects used widely varying information management strategies. Some subjects used elaborate hierarchical filing schemes whereas other stored little information for later use, and thus used much simpler filing schemes.

Half the subjects had extensive hierarchical information structures in which they kept information used as part of ongoing work tasks, or information that was kept for later use. Others used very simple schemes. For instance one subject reported that he had created a simple filing scheme for his information, where he would organize the information in very few categories. The categories either had project names or carried the name of the source from which the information was obtained. Another subject used an extensive filing scheme. When asked about the amount of information in this system he had the system count the number of files. It turned out he had almost 450 documents in 44 categories. It was quite a surprise for the subject that he had this much information stored.

One subject printed out most electronic information and kept it in a paper-based archive instead of keeping it online (the electronic version was deleted after it was printed). Thus only little electronic information was archived by this subject. Information such as emails were also printed and kept together with information related to a project, a customer, or some other organizational unit. However, it was noted by the subject that the information almost never used once a task was completed and the information has been stored in the paper archived. As long as the information was active and used it was not in the archiving system

As an example of storage and retrieval an interesting finding was that bookmarks were only used to a limited extent by the subjects. Moreover the bookmark feature in webbrowsers was seldom used, as found in other studies [ABC98]. If used it was typically only for a fairly limited number of bookmarks. However, one subject had an extensive hierarchy of bookmarks, but reported that only a few of them were used on a regular basis. A subject that did not use bookmarks reported that he had given up using an archive of bookmarks at some point. Now he relied on finding his information by searching on the web with a search engine. It is noticeable that this subject also reported to have an extensive use of the web for finding information relevant for his everyday work activities. Nevertheless that did not make him use the bookmark feature even for the set of websites he would visit on a daily basis. It was perceived as not worth the effort storing the information, as typing the address or searching was perceived as a more efficient way of retrieving information. Thus this finding indicates the dilemma that people want to spend as little time storing information as possible. On the other hand retrieval should also be efficient. In this example the subject

found that the use of search and remembering and typing addresses was the most efficient approach.

Problems in classification of information and creation of filing scheme were reported. As mentioned above there are examples of subjects reporting to have giving up filing information. A subject made the comment that the problem of reorganizing his information was solved when switching to a new desktop machine. It would have taken days or weeks to go through the information on the old system, and thus only the information that was considered critical to the work was transferred to the new system. The subject realized that the solution that would be the easiest would be to simply start over and create a new information structure on the new machine.

Some related observations were done in the study. Some subjects reported to reorganize the information occasionally and clean up the information that was no longer needed. Thus they would periodically carry out a cleaning session reorganizing and discarding information. Other subjects did not carry out such cleaning sessions at all, and some subjects kept little information in an archive.

Some subjects used similar information structures across different information types. That is, similar categorizations of the personal information were used in different applications. For instance information related to projects were kept in similar structures in the email system and in the file system structure. Keeping such similar structures across applications boundaries means that the subjects were manually maintaining and updating the multiple structures. This means extra information management tasks in terms of repeating tasks such as creating categories, and reorganization of the information structures. The use of the same or similar information structures lead to extra administrative burden on the user, and may be an indication of a need for a unified approach to the organizing of information.

9.1.6 Reminding

As seen in Table 9.2 only three subject used explicit electronic to-do items, using a system specifically for that purpose. However, reminding was also supported by different approaches. A couple of subjects reported that they sometimes e-mailed themselves a reminder. The e-mail being available in their inbox helped them

being reminded, as they would check the e-mail inbox often. The subject that used his mobile phone for reminders reported that sometimes he would get an audible reminder from the phone. This was sometimes inconvenient, as he would switch it off, and then actually forget about the reminder. This underscores the findings of other studies [BN95] that reminding can be accomplished with different kinds of information but is useful when the information is left in a location that is frequently visited.

Although subjects tended not to use explicit software for reminding, it was found that they used other ad hoc methods for keeping information reminding them of tasks to carry out. As mentioned above one such ad hoc method is that subjects send e-mails to themselves in order to keep a reminder. Also subjects reported to leave information, documents, or notes in particular places in order to be reminded of tasks to carry out.

The lack of sufficient support for reminding in interactive tools was also observed by the number of *Post-It notes* that were used in the office environment, and in one case on the frame around the display on the computer monitors. Again it indicates the importance of keeping reminders in places that are frequently seen or visited.

9.1.7 Discussion

A couple of interesting examples of unintended use were observed during the interviews. These examples provided some indications of systems not providing sufficient support for the apparent information management needs and the tasks that are carried out. There were also indications that subjects tried to obtain unified information management by means of unintentional use of existing systems. That is, they had created information management practices, which allowed them to have a single access point to a subset of their information, and thus one way of managing and interacting with this information.

One example of unintended use is the storage of information other than emails in the email system as mentioned previously. In that case making the email system the primary tool for information management. Not surprisingly information was kept in the email inbox to serve as reminders of things to do. It was also that the email inbox provided both a place for communication as well as a place

to keep to-do items.

Besides to-do items some subjects also used the email system to store bookmarks and documents, which were relevant to their daily activities, or archived for later use. Subjects reported how they emailed information to themselves and then later stored the information (as an email) into categories in the email system. This provided means of archiving different information types for later use.

From an interaction perspective this approach provides a unified way of organizing information of different types. Instead of organizing information by type by means of different applications, a single system is used to keep and organized different types of information items. The interviews revealed information such as to-do items, bookmarks, and documents being stored together in email systems into the information structure created by the subjects. Thus the approach allowed the subjects to store information of different types into the same information structure.

Another approach that was observed in the interviews was the use of the hierarchical file system to store different types of information. A couple of subjects had created extensive information structures with categories and levels of sub-categories in which they stored their information. This information included traditional desktop documents, such as text documents, spreadsheets, presentations and other files. This was primarily for archiving purposes, which also meant that the subjects did not keep personal information such as emails, appointments, contacts, and to-do items in this structure. Keeping a specific document in a particular place in the information structure could sometimes serve as a reminder of things to do.

One subject used this structure for keeping some bookmarks related to specific projects that were worked on. But typically the content, in terms of documents and from the particular websites were kept instead of a bookmark to the website where the information was found. The subject commented that it was found necessary to keep a local copy in case the information was not available for retrieval later when needed.

The interviews did not provide information to make generalizations or solid conclusions about the observed approaches to information management. For instance it is not possible to conclude based on the available sample whether email

or other systems have become a preferred way of doing management of information over traditional approaches. Additionally there are individual preferences.

Nevertheless, the observations that such systems are used in this particular way provide some interesting indications that some subjects prefer a unified way of managing their information. However, it has to be noted that not all subjects used these unified approaches, as some simply managed their information in different subsets. For instance one subject indicated the preference for keeping and managing different types of information by means of different systems, as it was then clear where the information could be found. This indicates that the highly individual nature of personal information management must be considered.

Half the subjects had created approaches like the ones mentioned above. That is an approach that would allow them some level of a single unified way of organizing at least a subset of their personal information. Typically the information that was used often or as part of ongoing work activities would be organized together across information types. This type of information management approach could be characterized as being an information-oriented way of organizing information. In terms of interaction this means a noun-verb oriented approach, where the information is accessed and managed through an information structure and available tools applied to the information in the structure.

One subject stated that he preferred the existing (task) tool-oriented way of organizing information. This means verb-noun oriented interaction where a specific information tool is the starting point for using and managing information of a specific type. Also information of different types is separated as different tools handle them. The remaining subjects cannot be clearly characterized as one or the other, as they would use both approaches or keep little information or only use a limited number of different information types.

The actual work tasks carried out and use of the various information types could have implications for the strategies chosen by the subjects. The observations and information provided in the interviews were not sufficient to establish explanations of relations between work tasks and the preference of one information management approach over the other.

The small sample size, and socio-demographic characteristics of the subjects mean that the results cannot be generalized to the general populations. As mentioned above personal information management practices and behavior varies

widely. Even within socio-demographic groups. For instance this is indicated by the identification of both "messy" and "neat" office types among the participants in the experiment. Generally this wide individual diversity poses a set of challenges for the experimental design and studies of personal information management. Also for software tools that aim to provide a suitable support of personal information management activities. Therefore it is paramount to have a comprehensive understanding of these activities and the strategies used.

Even though research has been carried out in the area for the last decades, still a comprehensive understanding of the subject is needed [WTN00]. A clear understanding of everyday core user tasks is needed. This includes personal information management tasks, which are carried out by a considerable amount of users of interactive systems every day.

The visions of *the paperless office* have been around for decades, but so far it has not become a reality. It is recognized that it may happen in domains where a set of standard procedures and standard forms are circulated. Such forms and procedures may be translated to electronic form.

However, the results provide indications and suggestions why the area of personal information management is far from paperless yet. The results indicate that the support of personal information management in interactive systems is not comprehensive. Not all aspects of personal information management are supported efficiently by current approaches in interactive systems.

Second physical form factors influence the situations of use. For instance the carrying of an artifact itself may be an issue posing limits to its use. For instance the Palm handheld device was designed to support personal information management and to be portable (carried in a shirt pocket) and thus could support mobile situations of use [Ber00]. Also input and output capabilities may be limited, which means limits of its use and possible situations of use.

Some traditional paper-based solutions may simply be more efficient than current interactive systems in some situations of use as indicated by the observations. The individual must assess the shortcomings of both approaches. A co-existence of approaches continues to be the case even decades after the initial studies of personal information management, which aimed to identify implications for the design of interactive personal information management systems. It is still the case even though the last decades have provided and researched

different approaches to personal information management. In any case the need for efficient support of personal information management has increased with the increased use of electronic information and the quantities of that information, such as, e-mails.

Thus, in the domain of personal information management the vision of the paperless office is so far unfulfilled. Further research is needed to further understand personal information management and experiments with different approaches to personal information management *if* a vision of paperless personal information management is even going to be fulfilled.

9.2 Results: Pilot Usability Experiment

The second part of the experiment was a simple usability experiment with the research prototype. As described in the previous chapter it was based on the thinking-aloud methodology. Subjects were given a brief introduction to the research prototype, and then provided with a number of tasks to carry out. During the experiments they were encouraged to think-aloud, and the observer wrote down their utterances.

However, before discussion the results from the experiment, the research prototype is briefly commented. First of all the research prototype is considered a simple proof of concept implementation as previously mentioned. That is, it has been demonstrated that it is possible to create a personal information management system on the basis of the model presented. On top of the model is implemented a user interface that provides a unified way of navigation, managing, and using the information in the system. The different adapters to different existing information sources enable the system to provide a unified interface to disparate information types.

Altogether this leads to a system that demonstrates an attribute-oriented approach and a unified interface to the management and use of personal information. Thus the research prototype serves as an initial proof of the overall concept. As such the limitations of the research prototype means that it does not allow for extensive usability studies to be carried out.

9.2.1 Tasks and Information

Initially the subjects were given a short introduction to the research prototype system, and the key concepts of the model on which it is based. After the introduction to the research prototype the subjects were given a set of tasks, see Appendix B and C.

- Check the email inbox
- Associate an email to a particular project
- Create email
- Send email
- Navigate the bookmarks
- Find and access a bookmark
- Navigate the music in the system
- Start a music track

As the prototype system contains adapters to existing information source it was possible to populate the system with different types of existing information. Additionally a set of default views and actions were made available in the system, thus the research prototype was able to simulate an environment for managing personal information. This included simple support for information such as e-mails, to-do items, calendar information, bookmarks, and music management and use. As mentioned in Section 8.3.2 during the experiment the research prototype system held approximately 500 information entities representing different types of personal information. Subjects were able to navigate, manage, and use this information during the experiment, as was it their own personal information.

As discussed in the previous chapter the purpose of the experiment was not to test the user interface of the research prototype as such. Rather the purpose of the tasks was to find out if the subjects were capable of understanding the underlying model on which the system is built.

9.2.2 Findings

The main result was that subjects were able to understand the simple tasks they were given and generally they were capable of understanding the conceptual system and use the actual research prototype. The immediate impression of the research prototype user interface was that it is not that different from user interfaces of existing systems, such as email systems. Not until starting solving the tasks involving association of attributes with information entities did the subjects observe the differences from existing systems. The interaction in the graphical user interface is similar to that of conventional systems.

Generally the subjects were able to navigate the information in the system, that is, selecting different views of information in the system. They navigated the views selecting different information entities and attributes associated with the entities.

Thus it seemed they were capable of understanding the concept, the model, and the parts of the model (entities, attributes, views, and actions). There were no major problems solving the tasks, which may be ascribed to the fact that the subjects were fairly experienced computer users. Nevertheless a couple of the subjects did have doubts on how to accomplish one task (association of a new attribute to an information entities), but after a while the observer gave brief hints to the menu in which to look, and the subjects were able to complete the tasks without further help.

Another issue where that the subjects had to realize that they could enter and change attributes directly in a view. This was not obvious for all subjects. From a user interface perspective this approach in the research prototype is not similar to the way in which information is entered or changed in existing systems. Such systems often require opening a different window and altering the information and then saving the changes. However, as soon as the subject realized they could alter information directly in the views it was no problem using this part of the system.

The experiment indicated that the part of the user interface that implements the assignment of new attributes to information entities could be improved. As mentioned above assignment of attributes in views were not a problem once understood. However, when an entirely new attribute needs to be assigned this

is done from the menu instead. Some subject tried if they could invoke a popup menu on a specific information entity. Although it could seem an obvious way of accessing such functionality it is not available in the current research prototype. Instead the assignment has to be invoked from the menu bar on top of the screen and then the attribute is assigned in a popup display that provides information about the specific information entity.

9.2.3 Debriefing Session

In addition to the tasks providing some initial results, the concluding debriefing session provided additional results and input for further development of the concept and research prototype. As mentioned the subjects were capable of understanding the concepts and use the research prototype to carry out the simple tasks. However, several of the subjects reflected on the system during the concluding debriefing session.

Surprisingly several of the subjects asked about the sharing capabilities in the system. They spontaneously asked about this, that is, the interviewer did not bring up the issue during the debriefing sessions or before. Several of the subjects reported that sharing of their information was very important for their jobs. Thus if this system did not support sharing of information they would not be able to use it.

Most of the subjects were interested in the approach. One commented that he -sometimes did not know in which application to look for some information that he knew he had in the system somewhere. Thus he would have liked to try out the approach more had it been a real system.

The discussion of generic vs. specific support was also brought up during the debriefing sessions. One subject commented that he would like to have a system like this, as he always had the problem of organizing his information into different places. Thus the subject seemed to like the idea of a unified approach to personal information management. However, in a debriefing session with another subject it was commented that he understood the concept, but that he would prefer that his information be divided in different applications so that he knew where the information was. This subject clearly preferred traditional specific support of personal information management. However, this approach leaves

the flexibility to organize information in a specific way or across information types. Not surprisingly this was not apparent to the subject after the short introduction to and use of the research prototype.

Finally, very much to the point a subject commented that he liked the approach and he could see some advantages, however, he would have to use the system for a longer time to find out if he would prefer to organize his information in this way.

9.2.4 Limitation

The experiments have been carried out as controlled laboratory experiments. The proposed model implemented in the research prototype could not be evaluated in a *real life* setting, which would be needed to be able to put forward solid conclusions about the NEXUS approach contrasted against other approaches.

The research prototype used in the experiments is experimental software with a set of limitations, as mentioned in Section 7.5. As such it does not represent a fully functional piece of software for personal information management. This poses a set of limitations on the experiments that can be carried out using the prototype. For instance the prototype cannot be used in its current state for extensive studies in a real work setting.

Future work would be to carry out experiments with real users carrying out their actual tasks using the system, and in a real work environment and in context of their actual work. Moreover it would be preferable to carry out such experiments over extended periods of time to make sure that the experiments are comprehensive and capture results from the entire system model. For instance it is rather difficult to test the value of the reminding feature, as testing this feature in a laboratory experiment tends to become somewhat artificial. It leaves the test subjects to speculate or guess if this feature would be useful to them in actual use. Thus, conclusions based on such results must be very cautious and only provide indications of its usefulness.

In order to carry out extensive research experiments the research prototype would have to undergo extensive software engineering, similar to the development and engineering of an actual software product. This would enable experiments with *real use* of the model. For instance it would enable experiments with differ-

ent quantities of information thus addressing the scalability of the model even further. However, this would require quite substantial engineering and development working hours. This dilemma seems to be a general challenge to research and experiments with novel ideas such as these. The question is to what extent a research prototype needs to be implemented to be able to carry out experiments that can provide results on which solid conclusions can be made.

Nevertheless, the results from the experiments are still valuable indications of the possibilities for personal information management software based on the attribute-oriented approach. In addition the experiments and results provide a set of suggestions for future research, which is also considered a contribution to the research area.

9.2.5 Discussion

The pilot usability experiment has provided some initial indications of the possibilities of the model as a foundation for attribute-oriented systems supporting personal information management. Additionally it has provided input for further development of the research prototype. An extended research prototype would potentially enable more comprehensive studies of the implications of this model and system for personal information management to be carried out.

The research prototype serves as a proof-of-concept, which indicate that it may support different levels of organization of information: Unstructured, semi-structured and formally structured information. However, the usefulness of the model in real-life personal information management situations is inconclusive based on the current results. The laboratory-based experiments does not provide sufficient basis on which to draw solid conclusions. Further development of the research prototype would be needed combined with additional and more comprehensive experiments compared to the pilot experiment described here.

Nevertheless, the experiments have provided some initial indications that the approach is worth researching further. The subjects were generally able to understand and use the model as the basis for carrying out simple personal information management tasks. Additionally most subjects provided positive feedback, especially during the debriefing sessions. Their reaction seemed to indicate at least a few of the subjects would possibly be interested in using the

research prototype system if developed further. This could potentially enable more comprehensive studies of the approach to be carried out.

Also, as no real information was used in the experiment it is inconclusive whether the approach will support reminding in an efficient way. However, the model and system enables leaving information in specific locations, which have been found in other studies to enable the support of reminding. This indicates that the model and system will at least support reminding at the same level as current systems.

An interesting topic that was brought up during the debriefing session was user preferences for task specific versus generic applications. It was clear that some subjects preferred one to the other. The one subject that clearly indicated preference for task specific approach said that he preferred the information to be in different places so that he knew different items were located. This utterance indicated that he found the generic approach in NEXUS too complex for his needs. This raises an interesting question whether such a generic approach leads to less or added complexity in a system for the support of personal information management. As mentioned the flexibility of the approach allows both kinds of information organization – task-oriented, type-oriented, or a combination. That is, if the subject preferred to keep his information in separate places, e.g. email in one view, and bookmarks in another, the approach allows this kind of organization. It is optional to organize information of different types together in the same views.

Moreover in traditional systems users have to learn different command sets for each application that they are using. Additional they must learn the specific user interfaces and ways of navigating the information in the different parts of the system. In this approach simplicity is provided through a limited and manageable set of commands. Users basically have to learn one set of commands that apply for navigation and management of information across the entire system. The core system elements that need to be learned in order to use the entire system are fever than in a traditional systems. Moreover it avoids the modal approach seen in traditional approaches.

Nevertheless it is questionable whether users will perceive the system as having a higher level of complexity compared to traditional approaches. The core facilities apply across the entire system, which means that the user needs to

reuse and apply the facilities in different situations for the support of different personal information management activities. Thus the approach may be perceived as complex, as users would have to figure out how to accomplish specific tasks using the generic tools, commands, and facilities provided by the system. This means that they would have to combine the core building blocks of the system and use them in combination in order to provide the support needed for an actual task at hand.

It could be argued that this issue could be addressed by providing a set of templates, default views, and commands in the NEXUS system. These facilities could guide the user to possible usage of the system in order to support specific work tasks. That is starting points to ways of supporting task, such as email. Nevertheless it is still unclear what this will mean to the perceived complexity of the system. Based on the data available from the study the implications of this generic approach to personal information management in terms of complexity cannot be determined. Further research would be required to get an understanding of this specific issue. However, there are some indications that subjects are divided in groups that have preference for the task-specific approach and groups that have preference for the generic approach. It is unclear to what extent this result is influenced by the current habits of the subjects. In any case this result underscores the level of individual preference and style in personal information management.

Another issue that is unclear is the scalability of the approach in terms of quantity of information and available commands in the system. The pilot system is working with 500 information entities, bus as no real use of the system has been carried out in the experiments the scalability issue is unclear. Further studies are needed in order to determine the possible scalability of the approach, possibly in real use situations of use, with increasing information quantities and structures evolving over time.

9.3 Research of Personal Information Management

In addition to the above-mentioned results this chapter is concluded with a discussion of some directions for further research. Personal information management is not well established as a research area. This project has mainly taken

a *human computer interaction* perspective on the phenomenon, but involving other disciplines as well. Other perspectives are relevant to involve in further research in the area.

9.3.1 Prior Research and Attention

Some initial studies of personal information management have been carried out since the early 1980's. However, in general personal information management has received limited research attention so far and does not have an established research community. Personal information management studies have primarily involved studies of specific tools, such as, email or bookmark management. Little attention has been on the fact that different types of personal information are used together, and therefore multiple tools are typically used together in order to fulfill everyday tasks and goals. Only recently some attention has been given to multiple information used together in personal information management [DCC+03].

Different studies have shown that personal information management is troublesome also with current interactive tools. The support of personal information management practices in current systems is limited. For instance much attention has been on archiving of information. Moreover as the quantity of electronically available information increases it becomes a difficult task to keep track of the information and organize it for later retrieval and use.

A set of novel approaches for the support of personal information management has been proposed [FFG96b] [DEL+00] [AKA99]. However, often they only address a subset of the problems involved in personal information management. Similar to the studies of personal information management practices these projects typically do no take an integrative or unified perspective on the problems involved in personal information management.

The simple study carried out in this project showed that subjects are using many different tools and artifacts in different situations in order to carry out their tasks and perform personal information management. This includes using the immediate environment as a placeholder for information, for instance to serve the reminding function. Information is spread across different places, locations, and organized by means of different schemes. In current interactive systems the

scheme is often determined by the user interface of the system rather than by the user. Information may also be spread across different appliances and supported by both paper-based and computer-based systems.

9.3.2 Relevance

The relevance of the establishment of further research in the area has several arguments. As mentioned studies have shown that personal information management is both difficult and time consuming. It is an activity that is carried out by millions of people (computer users) on a daily basis. Improvements in the tools and systems that support personal information management could thus have influence on many levels.

Thus one argument for further research in the area of personal information management is that our understanding of this complex phenomenon is still fairly limited. Another argument is that it could potentially have effect for millions of users that carry out personal information management to support their daily tasks and activities. Improved tools could potentially increase the efficiency and support of personal information management and user satisfaction. Improving support of the management tasks, could enable more focus on tasks at hand rather than management of the information itself, and thus improve individual productivity. Perhaps it could also be beneficial in a larger organizational and teamwork perspective if the individual has better personal information management tools.

It is relevant to establish appropriate methods and find directions for further research of personal information management. A set of problems can be addressed ranging from research and methodological problems, to specific problems in the current artifacts and tools used to perform personal information management. Further studies of personal information management practices should be carried out in order to develop a better understanding of the area. This includes a better understanding of human rationales behind personal information management practices.

9.3.3 Methodologies

Personal information management calls for interdisciplinary research due to its complexity. It involves human cognitive aspects, cognitive psychology, information retrieval, human-computer interaction, traditional computer science issues, and mobile systems and communication. There might be connections to other research disciplines and research areas as well.

Further work needs to be carried out to establish methodologies for studying personal information management. This includes methods of studying personal information management practices and use in existing environments, methods for the design of personal information management system, and methods for evaluating and potentially comparing systems that address personal information management. Methods used in prior studies of the area include:

- Interviews (semi-structured interviews about personal information management practices)
- Questionnaires
- Observations (of physical or virtual organization of personal information)
- Usability studies of novel approaches (have received little attention)

Some of the methodological problems were discussed in Section 8.1. Among these is the private nature of the information being studies. Additionally it involves studies of processes that are carried out over the course of days, weeks, and months in different situations. This potentially makes observational studies difficult to carry out in terms of data collection. Also longitudinal studies may be needed in order to study personal information management as a phenomenon, as information use and organization changes over time, and aspects such as reminding need to be studied. This pose a set of limitations on laboratory studies, as they cannot reveal all aspects of actual personal information management usage. This is also an issue when carrying out experiments based on prototype systems such as the one developed in this project.

Little work has been done in order to establish methods in order to evaluate and possibly compare different approaches to the support of personal information management. This could include methods that enable qualitative and

possibly quantitative ways to compare different approaches [WTN00]. However, the establishment of such methods is difficult due to the complexity involved in personal information management and the highly individual approaches and strategies used. The aim should be to evaluate an approach in general rather than specific features. Also, as mentioned previously preferably longitudinal studies should be carried out in order to capture use and changes over time. As the tools are used on a daily-basis such studies more or less rule out laboratory experiments.

9.4 Summary of Results

This chapter have presented and discussed results obtained from the two experiments that have been carried out. The experiments involved a semi-structured interviews and pilot usability experiments.

The semi-structured interviews have generally confirmed elements of the findings of prior research of personal information management [ABC98] [Col82] [Mal83] [Lan88]. This is not a surprise, as some of the issues mentioned in Section 9.1 have been identified in different studies of personal information management. However, it is noted that these studies were mainly carried out over a decade ago. Therefore it is interesting to see if those findings still apply considering the changes in interactive tools for the support of information management and increasing available information through computer networks. From the observations there are reasonable indications that the findings of previous studies are still valid. Even though new interactive tools have become available, the experiments suggest that it is still the same fundamental problems of personal information management that humans are facing.

It is therefore interesting to consider the reasons for these findings. Either it is due to fundamental human information management behavior. In the sense that no matter what tools is provided the underlying cognitive mechanisms and human information management behavior is the main reason for the problems described. Otherwise new interactive tools have simply not been capable of addressing these fundamental needs and provide mechanisms to support and facilitate the processes involved in personal information management.

Indications that would suggest the latter reason is the fact that even modern information management tools (as discussed in Chapter 4) require humans to organize information in hierarchical structures, which have been shown repeatedly as problematic in many situations. On the other hand novel solutions that aim to address the problems have not been successful on a large scale. The limited research results obtained in the personal information management area have not been applied successfully, thus little progress has been observed over the last decades. For instance hierarchical information organization still seems to be dominating in many interactive systems that involve personal information management. Examples are email software and bookmark management in webbrowsers. Recently novel approaches inspired by hypermedia-based systems have appeared as another alternative to the support of personal information management supporting structures in terms of interlinked information.

The main results of the pilot usability experiment are that subjects were capable of understanding the conceptual system and use the research prototype. They were capable of understanding the concept, the model, and the parts of the model (entities, attributes, views, and actions). Thus the experiment provided some initial indications of the potential of the attribute-oriented models for the support of personal information management based on the results of the pilot usability experiments involving the research prototype implementation. The model enables flexible information organization compared to systems mentioned above. However, further development and experiments would be needed to contrast the approach to traditional approaches to personal information management. This is also needed in order to establish solid results on the implications of the model and unified approach in terms of supporting personal information management.

Finally research of personal information management has been discussed as an additional result. The topic has received limited attention so far. Previous work has focused on specific information management and use. Methods for evaluating tools need to be established in order to potentially compare different approaches. Some suggestions for requirements for further research of the topic have been discussed.

Chapter 10

Related Work

The introduction (Chapter 1) includes a short summary of related work, and Chapter 4 includes a more detailed discussed of related work discussing existing system in terms of supporting personal information management. Therefore this chapter aims to briefly summarize the four projects that are considered closest related to this project. That is, to emphasize the similarities and differences among these projects contrasted against the NEXUS approach and the results.

The four approaches that are closest related to the approach studied in this project are considered:

- Lifestreams – using a temporal metaphor (electronic diary) as the basic organizational principle for documents. Organization is temporal and to some extent spatial. Also organization into substreams by topic can be created with search queries.
- Haystack – MIT research project. This topic-oriented organization of information is based on a set of W3C standards and specifically aims at the support of typical personal information management, such as emails, to-do items, calendar information, bookmarks, et cetera.

- Placeless documents – Xerox Parc research project (including the Presto file system). This project involves a framework and user interface approach for topic oriented organization of documents. Information items are visualized spatially in groups based on associated document properties.

- Lotus Agenda – Personal information management software from Lotus development (late 1980's). Enables topic-oriented organization of information items into different views based on the associated attributes. It is a text-based system that mainly uses tabular visualization of the information in the system.

These approaches can be characterized as attribute-oriented (semantic) approaches at some level. The approaches allow each information entity to be associated with descriptive attributes (metadata). However, there are only similarities in the underlying topic-oriented approach. As seen in the short overview above the visual representation of information entities with associated attributes, and the primary means of organization of these information entities varies among the approaches.

At the same time these approaches also represent novel approaches to the support of personal information management. Although the Placeless documents approach is aimed for general document management it has been applied for personal information management specifically.

10.1 Lifestreams

Although the Lifestreams model [FFG96b] can be characterized as an attribute-oriented approach it is distinctive being based on a temporal metaphor as the primary organizational principle for documents. The automatic organization of information by temporal attributes aims to address the problems introduced by hierarchical systems. A set of operations is available to facilitate finding information and organizing information into substreams based on search criteria. Therefore the approach allows other ways of performing information organization than temporal. Generally information is organized into one-dimensional streams, that is, substreams (lists) of information entities [Fre97].

The main focus is on the information organization and management problems, and less on actual data entry and use. Thus the model of interaction is still application-based similar to the desktop approach and in a sense replace the management of documents and files. In other words implementations of the Lifestreams model replaces the navigation layer of desktop systems, but not the actual data entry and manipulation. Meaning that there is a clear separation between the navigation and organization of information on the one side and data entry and use on the other. Information managed in the system has different types. Such types of information introduce modes, which is also apparent as different applications handle the different information types being handled by the system. This introduces a set of different user interface modes – similar to the compartmentalized desktop approach. Thus the system seen as a whole has a set of different modes of interaction. Contrary the NEXUS approach does not have the same separation of information management and data entry. Information and attributes can be entered directly in views, and such information associations affect the structure of the information. This is the primary way in which information is managed in the NEXUS approach. Organizing and manipulating information in the NEXUS approach is essentially the same, as it is supported through the same unified interface. However, it is noted that in NEXUS a similar issue could arise if the purpose was to replace the navigation and organization of coarse-grained information units.

10.2 Haystack

The Haystack project [AKA99] is also an attribute-oriented approach, which is specifically aimed at providing support for personal information management. Haystack is similar to NEXUS in that it experiments with "ways to incorporate the customization of user interfaces into the bigger problem of personalized information management by providing a platform upon which user interfaces can be modeled and manipulated with the same facility as other metadata" [HKQ02]. Contrary to the Lifestreams approach Haystack address both the issues of information management and data entry. The research prototype system built in the project is a system implemented from scratch. This means it implements both the information organization management facilities, as well as a user interface

specifically targeted to support traditional personal information, such as, email, calendar information, document handling, et cetera. Similar to the NEXUS approach Haystack enables the use of existing storage infrastructures and uses URLs to identify the entities (documents).

The main difference is in the details of the user interface. It appears that the Haystack research prototype has elements of an application-oriented user interface. That is, it has specific parts of the user interface dealing with specific information types, such as email and calendar. These parts may have special commands or particular user interface elements. However, it appears that the model on which Haystack is based essentially has the same level of flexibility as the one use in the NEXUS approach. The implementation of Haystack uses the Resource Description Framework (RDF) on which an implementation of NEXUS could possibly also be built. However, the emphasis in the NEXUS approach has been on personal information management aspects rather than the details of an implementation. Similar to NEXUS the Haystack project also aims to enable users to create custom views of the information in the system.

Haystack does not use the same approach to making commands (functionality) available in the system. In the NEXUS approach commands apply at the information levels rather than the view level. This means commands are modeless as they may apply to any information entity in the entire system regardless of the context. In traditional systems commands typically apply in a certain mode of operation involving specific information types. Thus commands in other systems may only apply in the context of a view of particular information types. However, as actions in NEXUS apply to information entities having certain attributes associated in practice the outcome may sometimes appear similar. Haystack does not appear to utilize a similar approach to commands.

10.3 Placeless Documents

The Placeless documents project [DELS99] is an attribute-based approach that tries to address the problems imposed by systems enforcing users to organize information into particular places. The Placeless Documents project has developed an architecture for storing documents based on their properties. The user as well as the system can specify these properties, and similar to NEXUS the system

allows arbitrary properties associated to objects and has a collection mechanism for aggregating documents. The Placeless approach enables the properties to serve as specification of access control and system backup. The properties can be shared enabling collaboration. These are issues that have been outside the scope of the current NEXUS research prototype implementation. However, this elegant way of utilizing information attributes for access control and backup is possible to add to an implementation of the NEXUS system as well. Also sharing of NEXUS information repository and views are considered obvious paths for future work on the approach and prototype.

Similar to NEXUS the Placeless architecture supports several existing storage infrastructures, and therefore enables the system to collect documents from multiple document repositories, such as, file systems, IMAP, HTTP, and databases. The systems also share similarities to the underlying model, however the information entities in Placeless Documents are *documents*, that is, text documents, spreadsheets, e-mail messages, et cetera. This leads to the same major difference as discussed in the Lifestreams approach. That is, the Placeless user interface mainly supports the *management* of documents in general from multiple repositories. Accessing, viewing and editing documents are done through different interfaces having different modes (similar to the compartmentalized desktop approach). Thus the most important difference between NEXUS and Placeless Documents is in the concept of unified interaction. Placeless Documents also aim to unify the interaction with multiple repositories of documents. However, Placeless Documents unifies the process of organizing, grouping, managing, controlling, and retrieving documents [DEL+00], but does not aim address unifying the process of information entry and editing. As mentioned above NEXUS aims to unify the interaction in the system further. That is, data entry and editing of the various information types is also unified, in that it is part of the model on which interaction is based. Additionally as commands apply to all information entities based on associated attributes.

The Placeless user interface combines the attribute-oriented approach with a spatial layout of information. The user interface presents groups of information entities visualized on a desktop-like surface of information. Thus it supports also a spatial way of organizing and retrieving information. Additionally it has an interesting feature – inclusion and exclusion lists. These lists enable the user to explicitly include entities in or exclude entities from a defined collection of infor-

mation entities. That is, explicit inclusion and exclusion of elements beyond the criteria that defines the collection (view). This approach introduces two inconsistent ways of defining the content of collections. The may be problematic from a usability point of view. Experiments would be needed to get an understanding of how users would understand and apply such inclusion and exclusion lists. It is possible that it may be difficult for the users to understand conceptually, as the criteria for the content in collections could be unclear. That is, there is a selection criterion, and then individual information entities are included and excluded from the collection. It could potentially be a source of confusion to the user that an information entity satisfies the selection criteria, but is still not included in the collection, because the entity is also a member of an exclusion list. Such experiments have not been carried out in the Placeless Documents project. NEXUS does not introduce such inclusion and exclusion lists, but it is possible to implement within the attribute-oriented framework.

10.4 Lotus Agenda

Lotus Agenda [KKB+90] is a system specifically aimed at supporting personal information management by means of an underlying attribute-oriented storage mechanism. Although Agenda is a text-based system only it shares some conceptual similarities with the NEXUS approach. It is also based on a simple model with a few basic components: Items, categories, and views that are used in the management and use of personal information. Categories (attributes) are slightly less flexible in Agenda, as they have a predetermined hierarchical structure. This is not a restriction in NEXUS, where attributes may or may not belong in an organization. However, the semantic analysis of information items available in Agenda is unique in that it automatically categorizes free-formed information entries, such as "Call Michael on Friday about the NEXUS project". Such an entity is automatically stored under multiple categories, such as phone calls, Michael, Friday, and NEXUS."

All personal information is integrated in the Agenda system. Similar to NEXUS views in Agenda may be different and in a sense represent a mode. However, the same commands apply to all items, independently of the attributes associated or the chosen view. As such like NEXUS Agenda does not have dif-

ferent modes of interaction in the same way as desktop systems and systems mentioned above. Also Agenda to some extent integrate the information management and data entry. Views of information entities also enable data entry similar to the way in which it is done in NEXUS. Compared to Agenda, NEXUS also has a unified way of making commands available in the system, as described in the previous sections. Like NEXUS, the Agenda system is also a system implemented from scratch with its own user interface facilitating both management and use of information. However, as it is a system from the late 1980's it obviously does not support multimedia content and does not have the same level of integration to other systems, as the other approaches discussed here.

Agenda is also similar to NEXUS in that information entities do not have a specific type. For instance an entity can be a both a note, and an appointment and a contact, whereas in the approach mentioned above that is not the case. In those systems the user has to decide in advance which type of information is entered into the system. NEXUS and Agenda also support free-formed information and for the information entities and information structures to evolve over time.

10.5 Summary

In this Chapter four different attribute-oriented approaches have been briefly discussed and contrasted against the NEXUS approach. The four approaches included are considered the ones that are closest related to the NEXUS approach.

The approaches are common in their use of an underlying data model, which allows for an attribute-oriented way of storing information. Although there are some minor differences at the data level they are not crucial for the support of personal information management. The main point is that the attribute-oriented approach leads to a more flexible way of supporting information management. Thus avoiding problems in existing systems, such as hierarchical storage and explicit naming of information entities.

The differences in the approaches are mainly found in the underlying understanding and assumptions about personal information management. Some projects mainly address the issues and problems of hierarchical storage of in-

	Unified information organization/management and interaction (generic)	Non-unified information organization/management and interaction (specific)
Generic task-support (modeless)	NEXUS Agenda Haystack	
Specific task-support (modes)		Lifestreams Placeless documents

Table 10.1: Comparison of different approaches to personal information management in terms of information organization/management and interaction (data entry and manipulation) and generic vs. specific task support

formation. The discussions of personal information management issues is this project have indicated a very complex area that involves a number of different issues and activities. Thus creating systems that provide support of personal information management demands considerable resources.

Personal information management is a complex area. A consequence of focusing on hierarchical information structures as the main problem in personal information management has been that some projects have focused on *replacing* this information organization principle by something else. The Lifestreams approach replaces the hierarchical organization by temporal-based organization to mention an example. There is no doubt that applications of temporal organized information exist and is useful for personal information management too. However, based on initial studies of personal information management it can be argued that the approach does not address and support the entire set of issues found in personal information management.

As pointed out in this project studies of personal information management have shown that different levels of information structure is relevant for different purposes and situations of use. Some level of hierarchical information organiza-

	L	H	P	A	N
Organization principle					
Temporal	x	x		x	x
Spatial		–	x		–
Attribute	x	x	x	x	x
Information					
Multiple view of same data	x	x	x	x	x
Information do not have explicit type				x	x
Unified interface					
Information management	x	x	x	x	x
Interaction data entry/manipulation				x	x

Table 10.2: Comparison of key elements of five different approaches to personal information management
L=Lifestreams, H=Haystack, P=Placeless doc., A=Agenda, and N=NEXUS
(x) indicates that the approach has the element, (–) indicates partly, and blank indicates that the element is not part of the approach

tion is useful in specific situations for personal information management also. For instance it may be used for archived information characterized by longer-term storage and retrieval typically based on some logical approach to the information organization. Thus, replacing this principle does not appear as an optimal solution for the support of personal information management. In the NEXUS approach the underlying model allows multiple ways of organizing information, with different levels of formality.

Finally, the approaches discussed in this chapter are different when it comes to the interaction level. Most projects have focused on the information management problems, and less attention has been given to the data entry and use of the information being organized. As mentioned above often there is a clear separation between the navigation and organization of information on the one side and data entry and use on the other. Such approaches often involve differ-

ent types of information leading to different modes of interaction in the system similar to the compartmentalized desktop approach.

The NEXUS approach does not have a separation between information management and data entry as such in the prototype user interface. Organizing and manipulating information in the NEXUS approach is essentially the same. That is, managing, viewing, manipulating, and using information is done through the same unified user interface. This enables the user to keep the focus on the information (content) rather than the organization of the information as such.

The way in which commands are made available in the NEXUS approach is different to the approaches discussed in this chapter. The commands are modeless as they may apply to any information entity and available attributes in the system independently of the views in which they are present. Traditionally commands in other systems only apply in the context of a view of particular information types or a specific part of the system.

The differences between the approaches discussed in this chapter are summarized in Table 10.1 and 10.2. It distinguishes between unified organization and non-unified information organization and interaction (data entry and manipulation). Also it distinguishes between generic vs. task-specific support, that is, modeless vs. mode-oriented approaches. Each of the different approaches has been mapped into this figure to indicate the focus of the approaches.

Chapter 11

Conclusions

A survey of prior studies of personal information management has been presented and initial attempts have been made to generalize some common findings reported from multiple studies of management of different types of personal information, such as, email use, document management, bookmark management, et cetera. Moreover an overview of different approaches to support personal information management in interactive systems has been discussed. This has pointed out problems in current approaches on different levels. These have been discussed from different perspectives, including cognitive perspective in terms of the role of memory in personal information management [Lan88]. This has lead to an identification of aspects that are relevant to consider in the design of personal information management systems in order to take into account the issues identified.

Common user interface problems in current personal information management systems have been discussed from a personal information management perspective. These issues have relevance for the construction of interactive systems aiming to support personal information management. A simple attribute-oriented model has been developed aimed at supporting issues in personal information management. Based on the attribute-oriented model a research prototype had been developed. It also provides a unified interface to personal information management. This means that the approach does not have a separation of information management and data entry. Additionally that commands in the

approach apply at the information levels rather than the view level, meaning that commands are modeless and apply to any information entity in the entire system.

Semi-structured interviews were carried out in order to get qualitative insights and an understanding of current personal information management practices and habits. Additionally to identify issues in current personal information management systems. The semi-structured interviews have generally confirmed findings from prior research of personal information management. As these studies were mainly carried out decades ago this study provides indications that even present tools do not provide improved support of the processes involved in personal information management. The experiments suggest that it is still the same fundamental problems of personal information management that humans are facing.

Additionally a pilot usability experiment involving the research prototype has been carried out. The pilot usability experiment indicated that subjects were capable of understanding the conceptual system and use the approach implemented in the prototype. They were capable of understanding the concept, the model, and the parts of the model (entities, attributes, views, and actions). Thus the experiment provided some initial indications of the potential of attribute-oriented models for the support of personal information management. Moreover the experiment provides initial indications that the approach provides support for different levels of structure and different levels of interaction with information.

Also research of personal information management has been discussed. The topic has not received a lot of attention and previous work has focused on specific information management and use. Methods for evaluating of tools needs to be established in order to compare different approaches. Some suggestions for requirements for further research of the topic have been discussed

The conclusion is that the studies provide some initial indications of the potential of attribute-oriented approach as the basis for system aiming to support and facilitated the process involved in personal information management. There are some initial indications that it may support different levels of information structure as well as information structures evolving over time. However, it is inconclusive whether reminding is better supported by this approach compared to existing approaches.

Further research of personal information management and approaches facilitating and supporting personal information management is needed. This includes further research of the implications of attribute-oriented approach on personal information management presented here. Further research is also needed in order to develop an understanding of user preferences for generic vs. task-oriented systems for the support of personal information management. Additionally research is needed to understand the implications in terms of complexity of the additional flexibility provided in this approach.

11.1 Contributions

The contributions of this thesis are considered to include:

- An attribute-oriented model on which personal information management systems may be built, enabling unified access and interaction with information in the system
- A research prototype implementation that serves as a proof-of-concept of the model for personal information management and a test bed for further experiments
- Experiment involving semi-structured interviews of subjects in order to get an understanding of present practices and problems in personal information management
- Pilot usability experiment with the research prototype implementation evaluating the approach in terms of providing support of personal information management
- A survey of prior personal information management studies and an identification and clarification of the issues involved in personal information management
- A comparative analysis of different approaches to the support of personal information management (in interactive systems)

- A set of directions for further research in the area of personal information management
- Identification of issues in personal information management that may inform designers of issues to consider in the design of such system

The project shares a set of similarities with four previous approaches to personal information management, Lifestreams [FFG96b], Haystack [AKA99], Placeless documents [DEL+00], and Lotus Agenda [KKB+90]. Although there are similarities, there are also important differences that are also considered contributions of this project. The key elements of the NEXUS approach includes:

- Integration of storage, management, retrieval, data entry and manipulation, in a unified user interface in the system. This means there is no separation as such between the navigation and organization of information on the one side and data entry and use on the other. This avoids different modes of interaction and the user may keep the primary focus of the information (content) rather than the organization of the information.
- Commands (functionality) is made available in a modeless way, as they apply to any information entity in the entire system independently of the views and context in which they are present. Traditionally commands in other systems only apply in the context of a particular part of the user interface or to particular information types.
- Provides generic support to personal information management in terms of the two above mentioned key elements, whereas other systems typically provide a compartmentalized task-specific support. The generic approach enables linking and aggregating of information having different types in more flexible ways than traditional hierarchical-based approaches.
- Supports multimedia information (compared to Agenda, whereas Lifestreams and Placeless Documents also support a variety of information types).
- Provides an open architecture that enable mapping of existing information types into the prototype implementation allowing further development and experiments to be carried out.

11.2 Future Work

Future work can be divided into four areas summarized below:

- **Further research problems**
  - Clarification of preferences for generic systems like NEXUS versus task-specific applications, as in the traditional desktop environments and personal information management software
  - Additional research in order to understand the implications of additional flexibility in terms of complexity in this approach.
  - Exploring the approach in a collaborative setting. For instance enable sharing of subsets of the personal information entities and study the implications of applying the model for information sharing and cooperative work, that is, applying the model beyond *personal* information management. This also implies a set of security issues.
  - This could also lead to possible applications of the model outside the domain of personal information management.
- **Further experiments**
  - Further quantitative and qualitative experiments of the approach, specifically to measure the capabilities of the system in terms of supporting personal information in longer terms (reminding and longer-term storage). Potentially carried out in a more realistic setting.
  - Scalability experiments, that is, include tests with different and larger quantities of information to get an understanding of the scalability of the approach (and possibly other approaches to personal information management).
  - Experiments with the model as basis for implementing personal information management support on different types of information appliances with different input and output capabilities – applying the approach in mobile systems.

  - Comparative studies of different approaches to personal information management in order to obtain an understanding of which aspects of the different systems provide support for particular aspects of personal information management and which aspects have less support.

- **Model and prototype**

  - Research prototype implementation improvements and optimizations.
  - Further refinements of the concepts, model, interaction techniques, and user interface improvements.
  - Continue the work on the model to further understand the implications for personal information management.

- **Technical issues**

  - Semantic analysis of information entities and data entry for automatic generation and association of attributes (metadata).
  - Formats for different user interfaces for different types of (mobile) information appliances.
  - Access to and integration of information in (possibly distributed) repository from multiple places, including distributed computing and network problems.
  - Repository scalability and indexing issues.
  - Repository sharing and security issues.

Additionally there are some interesting issues and questions where this project has not been able to provide a clarification.

The approach to personal information management support is based on a simple generic model and the user interface built on top also provides unified interaction. This can be characterized as a generic approach to personal information management. An interesting question is whether such a generic approach leads to less or added complexity in a system for the support of personal information management. In the discussion of the experimental results it was argued that simplicity is provided through a limited and manageable set of commands

and limited system elements that needs to be learned in order to use the system. These core facilities apply across the entire system, and the user should be able to reuse and apply the facilities in different situations and for the support of different activities.

However, on the other hand the approach may be perceived as complex, as users would have to figure out how to accomplish specific tasks using the generic tools and facilities provided in the system. This may not be obvious how these core building blocks of the system could be used in combination in order to provide the support needed for an actual task at hand. Nevertheless, providing a set of templates, default views, and commands could possibly improve this situation and guide the user to possible usage of the system in order to support specific work tasks.

Another issue is the scalability of the model and approach with increasing numbers of information entities and functionality in the system. In the experiments carried out in this project this was not an issue that was directly addressed. The research prototype held approximately 500 information entities in the repository when the simple pilot experiments were carried out. For the simple laboratory experiment it was simply assumed that this could be a realistic quantity of personal information in such a system. However, the semi-structured experiments indicated that some subjects kept quantities of information orders of magnitudes higher.

The underlying mechanism for information management and use is the attribute-oriented way of storing information and the command scheme for applying functionality to information entities in the system. It is not clear how these facilities scale with increasing numbers of information entities in the system, and increasing commands being made available in the system. Experiments are needed in order to get and understanding of the implications for storage and retrieval and generally information management practices when using the approach at different scales. Similar experiments could be interesting to carry out using the other approaches to personal information management enabling comparative studies to be carried out.

Bibliography

[ABC98] D. Abrams, R. Baecker, and M. Chignell. Information Archiving with Bookmarks: Personal Web Space Construction and Organization. *Conference proceedings on Human factors in computing systems CHI 98*, pages 41–48, April 1998.

[AG96] K. Arnold and J. Gosling. *The Java Programming Language*. Addison-Wesley, 1996.

[AKA99] E. Adar, D. Kargar, and L. A.Stein. Haystack: Per-user information environments. *Proceedings of the eighth international conference on Information and knowledge management*, pages 413–422, 1999.

[AM00] G. D. Abowd and E. D. Mynatt. Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1):29–58, March 2000.

[AOS+99] K. Arnold, B. O'Sullivan, R. W. Scheifler, J. Waldo, and A. Wollrath. *The Jini Specification*. Addison-Wesley, 1999.

[App01] Apple. Apple Mac OS X Cocoa Developer Documentation – Services Reference Documentation. 2001. http://developer.apple.com/techpubs/macosx/Cocoa/ProgrammingTopics/Services.pdf.

[BDHS03] V. Bellotti, N. Ducheneaut, M. Howard, and I. Smith. Taking Email to Task: The Design and Evaluation of a Task Management Centered Email Tool. *Proceedings of the conference on Human factors in computing systems*, (345–352), 2003.

[Ber00] E. Bergman. *Information Appliances and Beyond.* Morgan Kaufmann Publishers, 2000.

[BH94] B. B. Bederson and J. D. Hollan. Pad++: a zooming graphical interface for exploring alternate interface physics. *Proceedings of the 7th annual ACM symposium on User Interface software and technology*, pages 7–26, 1994.

[BN95] D. Barreau and B. A. Nardi. Finding and Reminding: File Organization from the Desktop. *ACM SIGCHI Bulletin*, 27(3):39–43, 1995.

[Boa01] R. Boardman. Category overlap between hierarchies in user workspace. *Interact*, 2001.

[BS00] V. Bellotti and I. Smith. Informing the design of an information management system with iterative fieldwork. *Conference proceedings on Designing interactive systems: processes, practices, methods, and techniques. New York City, New York, United States*, pages 227–237, 2000.

[Bus45] V. Bush. As we may think. *The Atlantic Monthly*, 176(1):101–108, July 1945.

[BYJR00] R. A. Baeza-Yates, T. Jones, and G. J. Rawlins. New approaches to information management: Attribute-centric data systems. *In Proceedings of the Seventh International Symposium on String Processing Information Retrieval (SPIRE'00), IEEE Comput. Soc*, pages 17–27, 2000.

[CFFG96] N. Carriero, S. Fertig, E. Freeman, and D. Gelernter. Lifestreams: Bigger than elvis. Techreport, Yale University, March 1996.

[CMS99] S. K. Card, J. D. Mackinley, and B. Shneiderman. *Readings in Information Visualization.* Morgen Kaufmann Publishers, 1999.

[Col82] I. Cole. Human aspects of office filing: Implications for the electronic office. *In Proceedings of the Humans Factors Society – 26th annual meeting*, pages 59–63, 1982.

[Cyp91] A. Cypher. EAGER: programming repetitive tasks by example. *Human factors in computing systems conference proceedings on Reaching through technology*, pages 33–39, 1991.

[DAW98a] A. K. Dey, G. D. Abowd, and A. Wood. CyberDesk: A Framework for Providing Self-Integrating Context-Aware Services. *Knowledge-Based Systems*, 11(1):3–13, September 1998.

[DAW98b] A. K. Dey, G. D. Abowd, and A. Wood. Cyberdesk: A framework for providing self-integrating context-aware services. *Proceedings of the international conference on intelligent user interfaces IUI'98*, pages 47–54, 1998.

[DCC+03] S. Dumais, E. Cutrell, JJ. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 72–79, 2003.

[DEL+00] P. Dourish, W. K. Edwards, A. LaMarca, J. Lamping, K. Petersen, M. Salisbury, D. B. Terry, and J. Thornton. Extending document management systems with user-specific active properties. *ACM Transactions on Information Systems (TOIS)*, 18(2):140–170, 2000.

[DELS99] P. Dourish, W. K. Edwards, A. LaMarca, and M. Salisbury. Using properties for uniform interaction in the Presto document system. *In Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 55–64, 1999.

[Der01] M. L. Dertouzos. *The Unfinished Revolution: human-centered computers and what they can do for us.* HarperCollins Publishers, 2001.

[EK95] M. Eysenck and M. Keane. *Cognitive Psychology.* Psychology Press, 1995.

[ELM+99] K. F. Eustice, T. J. Lehman, A. Morales, M. C. Munson, S. Edlund, and M. Guillen. A universal information appliance. *IBM Systems Journal*, 38(4):575–601, 1999.

[Fal96] J. Fallows. Organizational Behavior. *Inc Magazine*, March 15 1996. http://www.inc.com/magazine/19960315/1934-print.html.

[FF95] E. T. Freeman and S. J. Fertig. Lifestreams: Organizing your electronic life. *In AAAI Fall Symposium: AI Applications in Knowledge Navigation and Retrieval*, November 1995.

[FFG96a] S. Fertig, E. Freeman, and D. Gelernter. "Finding and Reminding" Reconsidered. *ACM SIGCHI Bulletin*, 28(1):66–69, 1996.

[FFG96b] S. Fertig, E. Freeman, and D. Gelernter. Lifestreams: An Alternative to the Desktop Metaphor. *In proceeding of CHI'96 Conference on Human Factors in Computing Systems*, pages 410–411, April 1996.

[FHH00] E. Frøkjær, M. Hertzum, and K. Hornbæk. Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? *ACM CHI Letters*, 2(1):345–352, April 2000.

[FLGD87] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. The vocabulary problem in human-system communication. *Communications of the ACM*, 30(11):964–971, 1987.

[Fre97] E. T. Freeman. *The Lifestreams Software Architecture*. PhD thesis, Yale University, May 1997.

[Gel85] D. Gelernter. Generative Communication in Linda. *ACM Transactions on Programming Languages and Systems*, 7(1), 1985.

[Gel94] D. Gelernter. The cyber-road not taken. *The Washinton Post*, April 1994.

[Gel98] D. Gelernter. *Machine Beauty*. Basic Books, 1998.

[GHJV95] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns – Elements of Reusable Object Oriented Software*. Addison-Wesley, 1995.

[GJSJ91] D. K. Gifford, P. Jouvelot, M. A. Sheldon, and J. W. O'Toole Jr. Semantic file systems. *Proceedings of the 13th ACM symposium on Operating systems principles*, pages 16–25, 1991.

[GN96] D. Gentner and J. Nielsen. The Anti-Mac Interface. *Communications of the ACM*, 39(8):70–82, August 1996.

[HKQ02] D. Huynh, D. Karger, and D. Quan. Haystack: A Platform for Creating, Organizing and Visualizing Information Using RDF. *Semantic Web Workshop, The Eleventh World Wide Web Conference 2002 (WWW2002). Honolulu, HI*, 2002. http://haystack.lcs.mit.edu/papers/sww02.pdf.

[JBD01] W. Jones, H. Bruce, and S. Dumais. Keeping found things found. *In Proceedings of Conference on Information and Knowledge Management*, pages 119–126, November 2001.

[JD86] W. P. Jones and S. T. Dumais. The spatial metaphor for user interfaces: experimental tests of reference by location versus name. *ACM Transactions on Information Systems (TOIS)*, 4(1):42–63, 1986.

[JN96] J. A. Johnson and B. A. Nardi. Creating Presentation Slides: A Study of User Preferences for Task-Specific vs Generic Application Software. *ACM Transactions on Computer-Human Interaction*, 3(1):38–65, 1996.

[Joh97] S. Johnson. *Interface Culture: How New Technology Transforms the way we create and communicate.* Harper Collins, 1997.

[Jør90] A. H. Jørgensen. Thinking-aloud in user interface design: a method promoting cognitive ergonomics. *Ergonomics*, 33(4):501–507, 1990.

[JRV+89] J. Johnson, T. Roberts, W. Verplank, D.C. Smith, C. Irby, M. Beard, and K. Mackey. The Xerox "Star": A Retrospective. *IEEE Computer*, 22(9):11–29, September 1989.

[JT97] S. R. Jones and P. J. Thomas. Empirical assessment of individuals' 'personal information management systems'. *Behaviour & Information Technology*, 16(3):158–160, 1997.

[KC82] J. F. Kelley and A. Chapanis. How professional persons keep their calendars: Implications for computerization. *Journal of Occupational Psychology*, 55:241–256, 1982.

[KKB+90] S. J. Kaplan, M. D. Kapor, E. J. Belove, R. A. Landsman, and T. R. Drake. Agenda: A Personal Information Manager. *Communications of the ACM*, 33(7):105–116, July 1990.

[Kwa89] B. Kwasnik. How a personal document's intended use or purpose affects its classification in an office. *Proceedings of the 12th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 207–210, 1989.

[Lan86a] T. K. Landauer. How Much Do People Remember? Some Estimates of the Quantity of Learned Information in Long-term Memory. *Cognitive Science*, 10(4):477–493, 1986.

[Lan86b] M. Lansdale. The role of memory in personal information management. *ECCE 3. european conference on cognitive ergonomics*, 3:157–163, 1986.

[Lan88] M. Lansdale. The Psychology of personal information management. *Applied Ergonomics*, 19(1):55–66, 1988.

[Lie98] H. Lieberman. Integrating User Interface Agents with Conventional Applications. *ACM Conference on Intelligent User Interfaces'98*, pages 39–46, January 1998.

[LR02] J. E. Larsen and M. G. Rose. Eight fallacies of distributed user interfaces. *In Proceedings of the Second Danish Human-Computer Interaction Research Symposium, Copenhagen, Denmark*, pages 38–39, 7th November 2002.

[LS02] J. E. Larsen and J. H. Spring. GLOBE: a highly available and scalable tuplespace. *In Proceedings SPIE ITCOM Conference on Java/Jini Technologies and High-Performance Pervasive Computing*, 4863:1–15, August 2002.

[Mac88] W. E. Mackay. More than just a communication system: diversity in the use of electronic mail. *Proceedings of the conference on Computer-supported cooperative work*, pages 344–353, 1988.

[Mal83] T. W. Malone. How Do People Organize Their Desks? Implications for the Design of Office Information Systems. *ACM Transactions on Office Information Systems*, 1(1):99–112, January 1983.

[MHP00] B. Myers, S. E. Hudson, and R. Pausch. Past, present, and future of user interface software tools. *ACM Transactions on Computer-Human Interaction*, 7(1):3–28, 2000.

[Mil56] G. A. Miller. The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *The Psychological Review*, 63(2):81–97, 1956.

[Mil68] R. B. Miller. Response time in man-computer conversational transactions. *Proc. AFIPS Fall Joint Computer Conference*, 33:267–277, 1968.

[Mil81] D. P. Miller. The Depth/Breadth Tradeoff in Hierarchical Computer Menus. *Proceedings of the Human Factors Society 25th Annual Meeting*, pages 296–300, 1981.

[MM03] F. Manola and E. Miller. Resource description framework primer. 2003. http://www.w3.org/TR/2003/PR-rdf-primer-20031215/.

[MP92] L. Meertens and S. Pemberton. The Ergonomics of Computer Interfaces. Techreport CS-R9258, Centrum voor Wiskunde en Informatica, CWI, Amsterdam, The Netherlands, 1992.

[MR00] S. H. Maes and T. V. Raman. Multi-modal interaction in the age of information appliances. *IEEE International Conference on Multimedia and Expo*, 1:15–18, 2000.

[MS95] K. Mullet and D. Sano. *Designing Visual Interfaces.* SunSoft Press, 1995.

[MSW92] R. Mander, G. Solomon, and Y. Y. Wong. A 'Pile' Metaphor for Supporting Casual Organization of Information. *Conference proceedings on Human factors in computing systems CHI'92*, pages 627–634, May 1992.

[NB97] B. Nardi and D. Barreau. "Finding and Reminding" Revisited: Appropriate Metaphors for File Organization at the Desktop. *ACM SIGCHI Bulletin*, 29(1):76–78, 1997.

[Nel80] T. H. Nelson. *Computer lib.* the distributers, 1980.

[Nie89] J. Nielsen. *Coordinating User Interfaces for Consistency.* Academic Press, 1989.

[Nie93a] J. Nielsen. Non-Command User Interfaces. *Communications of the ACM*, 36(4):83–99, 1993.

[Nie93b] J. Nielsen. *Usability Engineering.* Academic Press, 1993.

[Nie95] J. Nielsen. *Multimedia and Hypertext: The Internet and Beyond.* Morgan Kaufmann, 1995.

[Nie96] J. Nielsen. The Death of File Systems. *useit.com essay*, February 1996. http://www.useit.com/papers/filedeath.html.

[NJ94] B. A. Nardi and J. A. Johnson. User preferences for task-specific vs. generic application software. *Proceedings of ACM CHI'94 Human Factors in Computing Systems*, pages 392–398, 1994.

[NL93] J. Nielsen and T. K. Landauer. A mathematical model of the finding of usability problems. *Proceedings of ACM INTERCHI'93 Conference Amsterdam, The Netherlands*, pages 206–213, April 1993.

[NL94] J. Nielsen and J. Levy. Measuring usability: Preference vs. performance. *Communications of the ACM*, 37(4):66–75, April 1994.

[NMBG86] J. Nielsen, R. L. Mack, K. B. Bergendorff, and N. L. Grischkowsky. Integrated Software Usage in the Professional Work Environment: Evidence from Questionnaires and Interviews. *Human Factors in Computing Systems*, pages 162–167, 1986.

[NMW98] B. A. Nardi, J. R. Miller, and D. J. Wright. Collaborative, Programmable Intelligent Agents. *Communications of the ACM*, 41(3):96–104, 1998.

[Nor81] D. A. Norman. The trouble with UNIX. *Datamation*, 27(12):139–150, November 1981.

[Nor98] D. A. Norman. *The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution.* The MIT Press, 1998.

[Pem92] S. Pemberton. The Views application environment. Techreport CS-R9257, Centrum voor Wiskunde en Informatica, CWI, Amsterdam, The Netherlands, 1992.

[PF93] K. Perlin and D. Fox. Pad: an alternative approach to the computer interface. *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 57–64, 1993.

[PK97] M. S. Pandit and S. Kalbag. The Selection Recognition Agent: Instant Access to Relevant Information and Operations. *Proceedings of the 1997 international conference on Intelligent user interfaces*, pages 47–52, 1997.

[Ras83] J. Rasmussen. Skill, Rules and Knowledge; Signals, Signs and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man and Cybernetics*, 13(3):257–266, 1983.

[Ras93] J. Raskin. Down with GUIs. *Wired*, 1(6):122, December 1993.

[Ras94] J. Raskin. Intuitive equals familiar. *Communications of the ACM*, 37(9):17–18, September 1994.

[Ras00] J. Raskin. *The Humane Interface: New Directions for Designing Interactive Systems.* Addison-Wesley, 2000.

[RCL+98] G. Robertson, M. Czerwinski, K. Larson, D. C. Robbins, D. Thiel, and M. van Dantzich. Data mountain: using spatial memory for document management. *Proceedings of the 11th annual ACM symposium on User interface software and technology, San Francisco, California, United States*, pages 153–162, 1998.

[Rek99] J. Rekimoto. Time-machine computing: A time-centric approach for the information environment. *Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 45–54, 1999.

[Ros00] M. Rose. Music in the home: Interfaces for music appliances. *Personal Technologies*, 4(1):45–53, 2000.

[Shn96] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Techreport CS-TR-3665, University of Maryland Computer Science Department, July 1996. http://www.cs.umd.edu/Library/TRs.

[Tan95] A. S. Tanenbaum. *Distributed Operating Systems.* Prentice-Hall, 1995.

[Tay99] A. G. Taylor. *The Organization of Information.* Libraries Unlimited, 1999.

[Tes81] L. Tesler. The Smalltalk Environment. *BYTE*, pages 90–147, August 1981.

[Tuf83] E. R. Tufte. *The Visual Display of Quantitative Information.* Graphics Press, 1983.

[Tuf90] E. R. Tufte. *Envisioning Information.* Graphics Press, 1990.

[Tuf97] E. R. Tufte. *Visual Explanations: Images and Quantities, Evidence and Narrative.* Graphics Press, 1997.

[Ude01] J. Udell. The Future of The File System. *BYTE*, May 2001.

[Wal99] J. Waldo. The Jini architecture for network-centric computing. *Communications of the ACM*, 42(7):76–82, 1999.

[War00] C. Ware. *Information Visualization Perception for Design.* Morgan Kaufmann Publishers, 2000.

[Wei91] M. Weiser. The Computer for the 21st Century. *Scientific American*, 265(3):94–104, September 1991.

[WH01] S. Whittaker and J. Hirschberg. The Character, Value and Management of Personal Paper Archives. *ACM Transactions on Computer-Human Interaction*, 8(2):150–170, June 2001.

[WS96] S. Whittaker and C. Sidner. Email overload: Exploring personal information management of email. *Conference proceedings on Human factors in computing systems*, pages 276–283, April 1996.

[WTN00] S. Whittaker, L. Terveen, and B. A. Nardi. Let's stop pushing the envelope and start addressing it: a reference task agenda for HCI. *Human-Computer Interaction*, 15(2/3):75–106, 2000.

Appendix A

Guide for Semi-structured Interviews

As the experiment was carried out as semi-structured interviews the questions below where only used as a simple guide providing cues for the interviewer. Thus the interview was not necessarily carried out in that order and the questions were not explicitly asked if covered otherwise during the interview. The aim was open-ended questions to get the subjects to tell as much as possible about their information management practices and habits.

Test subject: #

Give me a tour of your office? Which information do you use on a daily basis?

For what purposes is the information kept?

How and where do you put your information? And why?

What are the methods used? And why?

Which tools are used to organize the information?

Appointments and to-do items

How do you remember appointments? Written down? Is a calendar used?

Do you use lists of things to do? Does it serve the purpose?

Keep notes (e.g. Post-It notes) as reminders?

Do you sometimes forget appointments or things that you need to do?

E-mail

How do you organize your e-mails?

Which mail program do you use?

How many e-mail is in your current inbox? How many categories?

How often do you use emails that has been archived?

Bookmarks

How do you organize your bookmarks?

Which browser used?

How many bookmarks are kept?

How many categories?

How often are these bookmarks used?

General questions

Does it occur that you cannot find the information that you are looking for?

What is the reason(s) if you cannot find information?

What are the problems involved in organizing your information?

Are there any problems in the way you organize your information (keeping, finding)?

Have you considered different approaches to organizing your information? Using different tools?

Information environment observed

Messy/neat?

Piles?

Tools used:

PC:

PDA:

Mobile phone:

Other:

Notebook:

Diary:

Calendar:

Other:

Paper archive:

Black/whiteboard:

Shelves:

Other:

Evaluation

How structured would you characterize your approach to organizing your personal information (scale 1-5, 1 = very good, 2 = good, 3 = neutral, 4 = poor, 5 = very poor).

Appendix B

Introduction to Pilot Usability Experiment

This is the short introduction to the pilot usability experiment provided to test subjects at the beginning of the experiment. Subjects were able to ask questions during and after the introduction.

Introduction

We are experimenting with new tools and ways in which to organize electronic information, such as, e-mail, bookmarks, music, calendar appointments, et cetera. That is, primarily electronic information that is kept in a machine, such as, a computer, a handheld device, mobile phone, et cetera.

A simple research prototype has been developed and it is used for simple experiments. The system can be used to organize information in a different way that is typically done in existing system. This provides new possibilities, and we are trying to find out if this approach provides good support of the ways in which we organize information.

If you feel uncomfortable proceeding with the experiment at any point during the session, you can just say so, and we will immediately stop the experiment. If you chose to do so the data collected will be destroyed.

Introduction to the system

This system – NEXUS – is a so-called personal information management application. It enables you to keep and organize your personal information, such as, your addresses, your calendar appointment, your e-mails, bookmark to webpages, organizing and listening to your favorite music, et cetera.

The system is built on three concepts:

- Entities – your information, such as, appointments, contacts, e-mails, bookmarks, etc.
- View – different collections of your information entities
- Attributes – information that can be attached to entities and allows you to organize your information
- Actions – commands that let you do different things with your information

On this system an experimental version of the NEXUS system is running, and what you see on the display is the main user interface of the system.

Appendix C

Simulated Information Management Tasks

In order to perform experiments with the research prototype and evaluate the model for personal information management, the test subjects were asked to perform a set of simulated tasks. While performing the simulated tasks the test subjects were "thinking aloud" and were also being observed. The simulated tasks were chosen to mimic typical everyday information management tasks. As such the subjects should be familiar with these at the task level, and the experiment is indicates how these work task are carried out using the research prototype.

The tasks are:

1. Find and browse the view containing the email inbox
2. Associate the email that has subject "NEXUS status" to the "nexus" project
3. Create an email with subject "status" and recipient "Jakob Eg Larsen"
4. Send the email you just created, and then find the email you just sent in the view containing emails "sent"
5. Find and browse the view containing bookmarks

6. Find and open the bookmark for the CTI homepage
7. Find and browse the view containing music tracks
8. Start a music track of your own choice. Then stop the playback of the track again.

Appendix D

Debriefing Statement

We are researching systems that aim to provide better support of personal information management. This work is based on how people do personal information management in traditional offices and using computer-based personal information management system. There are a number of problems in current systems, and our research aim to address some of these issues. For instance many people have trouble using the hierarchy folder structures present in current systems. One of the problems is that it is often hard to figure out where to file a certain document. Which folder should it go into – it could perhaps fit in more than one. That is also a problem when retrieving documents, as it can be hard to remember in which folder the documents are located, or the file name given a file. Especially if it has been a while since the document was used.

The research prototype that you have just used attempts to address some of these issues. It introduces different mechanisms for storing and retrieving information, which you have just used in simple tasks in the experiment. For instance instead of having a number of different applications to handle the different types of information that you use in traditional systems, this system enables the user to handle different types of information using the same system.

Therefore the results can be used to give indications if the approach is more suitable than other approaches to personal information management. We are interested in researching if this general system built on four simple concepts is more user-friendly, easier to learn, easier to use, and more efficient than tradi-

tional systems. As you may have noticed using e-mail, contacts, bookmarks and so on were handled in similar ways – using the four key concepts – in the entire system. So in a sense, in this system, doing your e-mail is not different from handling your contacts, your bookmarks, or any other personal information that you use.

Do you have any further questions or comments about the research or the research prototype system that you have used?

Do you have any concerns regarding the use of the data that we have collected during this session?

Thank you very much for participating – it has been a great help for the research.