Ergonomie: Mensch - Produkt - Arbeit - Systeme

Band 6

Pamela Ravasio

Personal Information Organisation

Studies on User-Appropriate Classification
and Retrieval Strategies and their Implications for
Information Management Systems Design

Shaker Verlag
Aachen 2005

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at http://dnb.ddb.de.

Zugl.: Zürich, ETH, Diss. Nr. 15579, 2004

Printed in Germany.

ISBN 3-8322-3560-4
ISSN 1610-1898

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen
Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9
Internet: www.shaker.de • eMail: info@shaker.de

Seeking to know is too often learning to doubt.
– *Antoinette Deshoulieres* –

Always remember you're unique –
just like everyone else.
– *Unknown* –

Preface

The achievement reported in this book relate to demands on information handling, categorisation, filing and retrieval on the personal computer, also called personal information management systems (PIMS).

With the rapid growth of computer storage-space, computer users have started to use their own computer to store on a large scale information from various sources and applications. As a result, computer users often experience that they need to re-access a particular previously filed document. They know that the document in question is located somewhere on their hard-disk, but they may have forgotten everything about it but what it roughly contained with respect to a specific context. For personal computers, this is a typical scenario where new and more flexible search mechanisms at the personal computer are needed.

Current desktop systems do not offer flexible search options when the search target is to a large degree undefined. These days, PC users have some few options for data retrieval, but these are closely linked to the hierarchical file structure of today's computers. In the future, more flexible and user oriented information handling, search, and retrieval interfaces for the personal computer need to be designed in order to enable more flexible and more intuitive search and human-machine retrieval interfaces. Such a redesign must go beyond mere design issues of the graphical user interface (GUI) only, as flexible information handling on the GUI level must be prepared trough the deeper, low-level file organisation.

This book also deals with another issue, which is not merely related to the design of PIMS, but actually applicable to it: the question of how to perform explorative research when the issue of exploration is to come up with or to define a target state rather than to test the implications of something already existing. In the presented context, the functionality of a PIMS-interface is to be defined. The traditional research approach in science for such goals is the empirical one based on the hypothetic deductive method. This approach is however not applicable to explorative research aiming at defining the qualities, demands and characteristics of something not yet known. A qualitative research approach meets the need of explorative research and offers a structured and systematic way to produce qualitative data. Much care has therefore been taken in order to describe the qualitative *experimental* research method in a way accessible to the general audience of researchers in the field of Human-Computer Interaction (HCI).

In this book, the author presents thorough work based on results that will serve as basis for the development of intuitive and more flexible personal information management, search and retrieval systems. She has taken another approach than the mainstream in this field: instead of implementing a new interface based on the designers more or less intuitive belief in improvements, she has taken the effort to start with basic preparatory investigations. This required the determination of the actual status quo of technical development, the evaluation of the strengths and weaknesses of current systems, as well as a thorough investigation about what users need. The application of the qualitative experimental research method in this case,

implied furthermore large efforts in user data analysis (transcribing, analysing, interpreting, inferring). The results of this investigation are therefore a valuable contribution to the field of Human-Computer Interaction.

Zurich, November 2004

S. Guttormsen Schär H. Krueger

Acknowledgements

I would like to thank Prof. Dr. Dr. H. Krueger and the members, actual and former, of the Institute for Hygiene and Applied Physiology at the Swiss Federal Institute of Technology, Zurich, for the many opportunities to gain new insights into, to me, yet unknown areas.

Also, I would like to thank my MMI colleagues for the many interesting and instructive discussions and the laughs we shared.

Special thanks go to Vincent Tscherter for being a friend to share my laughs, my deceptions and many of my interests with. Without him and his tireless moral support, this thesis probably would not have been written at all.

And, last but not least, thanks go also to my parents, Rosa María Kebschull Rial and Ernst Ravasio, for their continuous support during my academic and non-academic life.

Pamela Ravasio

Table of Contents

Abstract

Personal information consists of pieces of information that individuals collect, store on their computer, and try to find again at some point in the future when context makes them relevant again.

Since the beginning of the 1990ies, the total amount of information stored globally, and the amount of personal information stored by individual users, has grown exponentially. Whilst 'good' organisation of information has always been a major concern for both the individual and systems designers, today 'efficient' and 'intuitive' retrieval of stored information is slowly gaining greater attention.

This is why the presented thesis aims at investigating issues related to the organisation and, even more importantly, the access to and retrieval of previously organised information, and the respective technical infrastructure.

The results then, can be categorised as follows:

Method In the context of HCI, the adaptation and definition of the *qualitative experiment* as an investigative method in its own right.
The qualitative experiment is derived from two methodologies: the inductive type of qualitative methods, and the deductive type of quantitative experimental methods. Therefore, the qualitative experiment aims to achieve two things: whilst being exploratory, it still follows a well-defined procedure trying to prevent accidental outcomes and results.

System *Dos and don'ts* for personal information management systems.
In guidelined, semi-structured interviews, the problems and practises of users when handling their information were addressed.
The range of results obtained cover details in the lower levels of a system - e.g. the file system - as well as nuanced aspects directly visible in the graphical user interface (GUI) itself.

Theory Definition of user-friendly, *context-independent metadata*, and *tactics and rules* actually employed by users for searching and classifying information suitable to the user.
When personal context can no longer be relied on, e.g. after long periods of time, or in cases of group memories (i.e. collaboratively used storage spaces), information must still be storable and accessible as intuitively as possible. User-friendly, context-independent metadata supports intuitive access to data

by the end-user even in situations where the original context of information is no longer available.

The results of this thesis are of general interest whenever information and document management systems are being designed, implemented or dispatched, and where information display and visualisation is an issue.

Kurzfassung

Persönliche Information besteht aus vielen einzelnen Informationsteilchen, welche von einer einzelnen Person gesammelt und auf deren Computer gespeichert werden, um wiederverwendet zu werden, sobald sie wieder relevant sind.

Global ist die Gesamtmenge der gespeicherten Information und daher auch die Menge, welche durch jeden Einzelnen gespeichert wird, seit dem Beginn der 1990er Jahre exponentiell gewachsen. Früher war das 'optimale' Archivieren (Klassifizieren) sowohl für Systemdesigner als auch für den einzelnen Anwender von eminenter Wichtigkeit. Heute kommt diese Bedeutung dem 'effizienten' und 'intuitiven' Wiederfinden von Information und dem Zugriff darauf zu.

Die vorgelegte Dissertation beschäftigt sich mit Fragen der Informationsorganisation, vor allem dem (Wieder-)Zugriff auf bereits abgelegte Information, sowie der technischen Infrastruktur, die der Informationsoganisation zugrunde liegt.

Die präsentierten Resultate können wie folgt zusammengefasst werden:

Methode Die Idee des *qualitativen Experiments* wurde aufgegriffen und für die Disziplin der Human-Computer Interaction (HCI) definiert.
Das qualitative Experiment führt Elemente zweier methodischer Seiten zusammen: Elemente qualitativer Forschungsmethoden induktiver Art, und Elemente quantitativ-experimenteller Methoden deduktiver Art. Das qualitative Experiment zielt darauf ab, explorativ zu sein und gleichzeitig einer gut definierten Prozedur zu folgen, damit zufällige Resultate und Fehler vermieden werden können.

System *Dos und don'ts* im Bereich der persönlichen Informationssysteme.
In richtlinienbasierten, semi-strukturierten Interviews wurden die Probleme und Gewohnheiten von Benutzern im Umgang mit ihrer Information analysiert.
Die Spanne der kristallisierten Resultate reicht von Details in den unteren Schichten eines Systems - z.B. im Dateisystem - bis hin zu Elementen, welche in der graphischen Benutzungsschnittstelle direkt sichtbar sind.

Theorie Wir definieren benutzungsfreundliche, *kontextunabhänginge Metadaten*, sowie *Taktiken und Regeln* welche vom Benutzer verwendet werden, um seine Information zu archivieren und wiederzufinden.
Information muss so intuitiv wie möglich gespeichert und (vor allem) wieder

auffindbar sein. Selbst wenn der Kontext, innerhalb welchem die Information gespeichert wurde, nicht mehr der Aktuelle ist. Eine solche Situation kann z.B. entstehen, wenn Speicherplätze kollaborativ oder lange nicht mehr genutzt werden.
Benutzungsfreundliche, kontextunabhängige Metadaten ermöglichen einen für den Endbenutzer intuitiven Zugriff, selbst in Situationen, wo der ursprüngliche Kontext eines Informationsteilchens nicht mehr existiert.

Die Resultate der vorgelegten Dissertation sind daher sowohl für das Design, die Implementation und die produktive Einführung von Informations- und Dokumenten-Managemensystemen von praktischer Relevanz, wie auch dort wo Informationsvisualisierung ein Thema ist.

Transfer to Practice

Summary

Information Systems have been an issue in research ever since their invention. Interestingly, it has been overlooked that desktop computers and the information each individual collects and stores to the file system, form another category of extremely small, but none the less potent 'personal information systems'.

This chapter hence addresses questions that are of interest wherever information is created, worked with and administrated by humans in their working context. Amongst these questions are: What general types of activities need to be supported by (personal) information management systems? Where do storage facilities fail and why? How would alternatives need to look like?

Introduction

Since their invention, Information Systems have been an issue in research. However, Information Systems research has mainly focused on large-scale databases predominantly used in the industrial context, and consequently, and which intended for *structured* storage of large amounts of often complex information.

It is usually overlooked that another category of small, but none the less potent 'personal information systems' conquered the individual office workers and their desks: The desktop computer and the information each individual collects and stores to its file system.

At first sight the difference between large-scale and personal information systems appear to be merely issues of size and power; however, a closer examination reveals a deeper distinction. Databases serve to store well-structured pieces of information, normally generated by well-defined processes, which are handled by experts. PCs on the other hand, serve to store rather unstructured information from all sorts of sources which are handled by an extremely broad user group, both in terms of skill and portfolio expertise.

With these differences in mind, it is astonishing to note that many of the paradigms used today for information handling, organisation and retrieval in PCs originated from large-scale databases and were adapted to the world of PCs (prominent examples in this context are the Windows search functionality, as well as the WinFS file system to be released with the Windows operating system by 2005/6).

Furthermore, modern desktop systems are designed to serve more than one master simultaneously. Besides information storage, they are *also* used for ad-hoc work practises like emailing and the maintenance of an electronic agenda, the creation of textual documents like letters and protocols, the execution of spreadsheet calculations etc. The storage and retrieval of information is therefore just one of many tasks that must be successfully accomplishable with these systems. Nevertheless, these are tasks where individuals potentially profit most from the mere existence of desktop systems as opposed to the pre-computing era.

Since the 1990s, the quantity of stored information has grown exponentially over time (Section 1.1). However, the means with which we administer this flood of information storage have remained virtually unchanged since the days when *personal* information storage was unheard of. Whilst *user-suitable information organisation* has always been discussed as a topic in optimisation, consideration of information access and retrieval has been limited.

Whilst a desktop system serves many masters, it is our opinion that the most crucial one is *accessing* information storage for re-use, since it is only then that collected information becomes ultimately useful again. Therefore, the primary focus of this thesis is 'good' retrieval without neglecting classification.

Work Practices

Working with desktop systems, or personal computers as they are called, encompasses two fundamentally different types of activities:

- Goal-driven activities.
  Examples are: the writing of a letter or protocol, designing a poster or calculating a months production throughput.

- Ad-hoc activities (often called 'knowledge work').
  Examples are note-taking, document commenting, and the organisation of the personal agenda.

While support of the former has been improved and optimised over the years through a range of specialised application programs, quality support for the latter is *still* extremely hard to come by.

As a consequence, there exists a tremendous potential for innovative development in this area which must be confronted in the near future.

Existing hierarchical storage facilities work fine in the context of 'goal-driven' activities; but are not so appropriate in the context of 'ad hoc' work practices. While it makes sense to compel the user to classify information for 'goal-driven' work, this constraint is not appropriate for 'ad hoc' work. Unfortunately, there is no way to prevent the act of classification, as conceptually practised in the physical world through means of piles and similar facilities.

The inequality in support for the two types of activities manifest further in the system's internals:

- Documents that are created or received through 'goal-driven' activities are under the user's command which means he or she can store them centrally to locations intended for his or her use. The storage location therefore does not depend primarily on the activity from which the information stems from, but is explicitly assigned by the individual user.

- However, pieces of information created and received via 'ad-hoc' activities are *spread* over a variety of locations in dependence of the activity they originate from, with the user having very little possibility to intervene. Emails, for instance, are assigned one specific location by the system, bookmarks a second and electronic sticky notes a third. This results in a lack of consistency and comprehensibility.

What can not be overseen anymore is the fact that people *do* use information archives. And, it is unlikely that those archives will entirely disappear because of both personal preferences and the fact that some businesses are legally required to keep them (e.g. accounting). Information contained in archives can be looked at and processed from at least three different angles:

1. *Task oriented*, i.e. focused on the task to be accomplished.

2. *Context oriented*, i.e. focused on other documents, programs and tasks at hand concurrently.

3. *Content oriented*, i.e. focused on the actual information encapsulated in a specific document.

An individual user will not exclusively use one approach to work with information, but is more likely to make use of all three, depending on a given scenario.

These three different ways of looking at, and working with, information, impact directly on information storage and retrieval strategies. We have seen that the existing rigid, hierarchical structure does not support one of three concepts successfully, let alone support all three concepts simultaneously.

The question is therefore: In the eyes of potential users, how should a storage facility 'work'?

The fundamental idea is that of a library catalogue. Instead of the typical 'librarian' fields (such as ISBN numbers and the like), to the average user more intuitive descriptors[1] would be used. Among such descriptors would be the document's type (e.g. letter, protocol) or a sketch of its first page.

These descriptors were the ones we were after. We were able to identify a set of them that seems very plausible also from a non-scientific point of view.

It is also interesting to note that whilst humans find these descriptors to be intuitive, logical and structured, (computer) technology perceives them as the opposite (i.e. unstructured) - because it is still a challenge to have the type "letter", for example, defined in a way that a computer can cope with.

Usability studies, prototyping and qualitative field methods have found a relatively wide application in industrial product design and evaluation.

However, in the commercial-industrial context, quantitative investigational methods are not normally applied. The reasons for this are many, but the main ones are probably due to the fact that lab experiments require a large amount of effort and a well tuned machine, and their degree of abstraction is often higher than makes sense for an applied industrial context.

While these were essentially the same problems as we faced during the work on this thesis, we were able to overcome this obstacle through the identification and refinement of a method originating from sociology and social psychology: the qualitative experiment.

It is a method that applies a structured technique and aims to *discover* relations and structures instead of assessing them. As a consequence, it is an interesting methodical option for practitioners that try to go beyond existing (methodical) limitations.

[1] Throughout this thesis, we call them "user-friendly metadata".

Comment and Conclusion

The race for a new metaphor that will ultimately solve all the problems inherent to the omnipresent desktop, was already launched a decade ago (see e.g. (Fertig et al. (1996b))).

Previously, the impetus for seeking a better metaphor for the PC's graphical user interface was its non-compliance with its physical counterpart, in the future the quest will be driven by the hunt for efficient working practices and procedures suitable to the user.

As a consequence, the new metaphor needs to be shaped by the 'optimal' working practices (instead of defining the optimal practices in dependence of the metaphor). Innovations always bring along changes. The challenge is to motivate changes that perceivably improve existing practices and do not force the users into adapting without giving them any clearly visible added value.

Whilst some of the current efforts are heading in the right direction (e.g. (WinFS (2004); Brain (2004)), it is likely that these efforts will ultimately fail as long as they try to conserve the existing desktop metaphor to such a great extent. These efforts also indicate that a change is coming in the not too distant future.

In technology, it can be assumed that "everything is possible" and what remains is the question of *how* to do it. It is therefore important that the focus of attention is turned towards this question of *how*, i.e. towards the users and the way they accomplish their work, in order to tune systems accordingly. This means that the user must claim the centre of consideration previously occupied by systems and technologies.