# Computer Support for
# Documentation Work

Morten Hertzum

Ph.D. thesis

Department of Computer Science, University of Copenhagen
Universitetsparken 1, 2100 Copenhagen Ø, Denmark

DIKU report 94/20
August, 1994

# Preface

This report is submitted in fulfillment of the requirements for the Ph.D. degree at DIKU, the Department of Computer Science at the University of Copenhagen. It concludes work done between January 1992 and August 1994. My Ph.D. study is documented in five articles, a technical report, and this report which is the summary. Three of the articles are published, the last two submitted for publication.

The study is about documentation work and how it is best supported by computers. With its concern for representing knowledge in recorded data, documentation work is central to intellectual activity and applies to all sorts of professional work, for example the work of lawyers, researchers, and civil servants. The study takes the actual performance of documentation work as its starting point, i.e. the approach is primarily empirical and experimental.

Many persons have contributed to my work, as coauthors, critics, interviewees and in various other ways. My indebtedness to them is specifically expressed in the acknowledgments of the single papers. Without repeating their names I like to take this opportunity to wholeheartedly repeat my thanks to all of them. My very special thanks to Erik Frøkjær, my Ph.D. advisor, for pushing me further than I would otherwise have gone, for numerous inspiring ideas and insights, and for his attention to detail whether in questions of planning, rewording, or life in general. Also, I want to thank the University of Copenhagen for the grant, *kandidatstipendium*, paying my salary and DIKU for providing good working conditions and travel funds. Finally, my warmest thanks to my family and friends who showed interest in my work or just recognized it as my odd obsession.

*Morten Hertzum*
*August, 1994*

“Books are not made to be believed,
but to be subjected to inquiry.”

(U. Eco, *The Name of the Rose*)

“Prefer the specific to the general, the definite
to the vague, the concrete to the abstract.”

(W. Strunk & E. B. White, *The Elements of Style*)

# Contents

## Appendixes

# 1. Introduction

Documentation work is an activity known from within by countless professionals such as researchers, lawyers, and civil servants. Though the basic objective of their documentation work is the same different professionals perform it differently depending on the kind of organization they are employed in, the nature of their work, their personal preferences and so on. In studying and supporting documentation work it is essential that this variety is taken into account. The common core of the documentation work is obviously important, but to identify it and put it to use in a documentation system require close attention to the specific characteristics of the situation for which the system is developed.

The basic meaning of the term *documentation work* is apparent from its etymology. The words *document* and *documentation* both derive from the Latin `documentum` which means lesson, example, proof (`docu-` in turn derives from `docere`, to teach). *Document* denotes something preserved and serving as evidence, usually in printed or written form. Though *document* and *text* are often used interchangeably the word *text* has quite different origins. *Text*, like textile, derives from the Latin `texere` which means to weave. Thus, *text* is interwoven words, a context-free description which says nothing about why the text is written. *Document*, on the other hand, presupposes the notion of time: A document is a text related to time by being produced and preserved at some point in time to enable retrieval and use at a later point in time. The documents, references, and other material preserved or retrieved in a given situation is referred to as *documentation*, as is the process of preserving and retrieving them. *Documentation work* denotes this last, work process sense of *documentation*.

Often, documentation work is thought of as a rather mechanical or clerical activity contrasted to the primary work which is characterized by words such as *intellectual* and *thinking*. In 1960 this distinction led Licklider (1960) to believe that computers would be able to improve or facilitate thinking and problem-solving. To justify his belief Licklider offered an informal study of his own work process, a study which distinctly illustrates the extent of his documentation work and how closely it is intertwined with his primary work:

> “It soon became apparent that the main thing I did was to keep records, and the project would have become an infinite regress if the keeping of records had been carried through in the detail envisaged in the initial plan. It was not. Nevertheless,

I obtained a picture of my activities that gave me pause. Perhaps my spectrum is not typical—I hope it is not, but I fear it is.

About 85 per cent of my "thinking" time was spent getting into a position to think, to make a decision, to learn something I needed to know. Much more time went into finding or obtaining information than into digesting it. Hours went into the plotting of graphs, and other hours into instructing an assistant how to plot. When the graphs were finished, the relations were obvious at once, but the plotting had to be done in order to make them so. At one point, it was necessary to compare six experimental determinations of a function relating speech-intelligibility to speech-to-noise ratio. No two experimenters had used the same definition or measure of speech-to-noise ratio. Several hours of calculating were required to get the data into comparable form. When they were in comparable form, it took only a few seconds to determine what I needed to know.

Throughout the period I examined, in short, my "thinking" time was devoted mainly to activities that were essentially clerical or mechanical: searching, calculating, plotting, transforming, determining the logical or dynamic consequences of a set of assumptions or hypotheses, preparing the way for a decision or insight. Moreover, my choices of what to attempt and what not to attempt were determined to an embarrassingly great extent by considerations of clerical feasibility, not intellectual capability."

During the past 30 years, computers have been assigned a key role in the various efforts to support documentation work. Computer support for documentation work is a central and much studied issue in a broad range of literature including computer-supported cooperative work, database technologies, human-computer interaction, hypertext, information retrieval, management information systems, and office automation. However, time and time again the development of satisfactory systems has proved, and still proves, more difficult than expected. The problem has not been properly understood yet, let alone solved; rather, it has turned out to be inherently difficult and involve manifold factors.

### *1.1 Field of study*

Computers are used by people to perform tasks. This seemingly innocent assertion identifies three elements fundamental to situations in which computers are used—people, computer-based tools, and tasks. However, their relationship is much more subtle than the assertion suggests. Naur (1965) emphasizes that the essence of the relationship is its symmetry, i.e. either of the three elements can be understood only in the context constituted by the two others. Thus, for example, tasks exist only insofar as they are recognized by persons and only relative to understood tools. Changing the task, obviously,

changes the opinion about what is a proper tool, but changing the tool also changes the opinion about what constitutes the task. The people-tools-tasks triangle is concerned with individuals and their perception, it does not address the cooperative aspects of work. Leavitt (1964) deals with organizations and organizational change, but except for an additional organizational element he is concerned with the same three fundamental elements as Naur. Leavitt's main point is that successful, organizational change of one of the elements will usually effect or require changes of the other elements as well. Thus, he too emphasizes the vital importance of the dynamic interplay between the elements, though at the organizational, rather than the personal, level.

This study concerns computer systems supporting professionals' documentation work in organizational settings. In terms of Leavitt's model the field of study is the situation constituted by the interplay between four main elements, see figure 1.1. These elements are described briefly below, followed by an introduction of the two relationships which are the key to the themes running through this study.

*Documentation work* is a continuous activity intertwined with the work to be documented. In its broad sense documentation work includes the production, filing, management, and retrieval of the material documenting the primary work. In its narrow sense the production of the documentation is considered part of the primary work, and documentation work is limited to the handling of the documentation—filing, managing, and retrieving it. In this study the term documentation work is used almost exclusively in its narrow sense.

*Professionals* are subject specialists in their task domain and can not be expected to possess special knowledge about documentation work or computers. Further, professionals are distinguished from clerks, i.e. persons employed to keep records, handle correspondence or the like. The professionals here studied include chemists, civil servants, lawyers, and a group of semi-professionals, computer science students.

*Text Storage and Retrieval Systems (TSARS)* are computer-based systems intended to support documentation work. Some TSARS provide access to fixed or externally updated collections of texts; others are directed towards handling documents internal to an organization. Thus, depending on the system and the privileges held by the user, TSARS may provide facilities for retrieval only or, say, for filing, management, and retrieval.

*Organization* is the structures and the rationale established to handle the cooperative aspects of work, such as coordination and delegation. Often, different persons are responsible for different parts of the documentation work and therefore experience it—and the TSARS—very differently. Also, the organizational element introduces a managerial perspective on documentation work, a perspective in which control is a central issue.

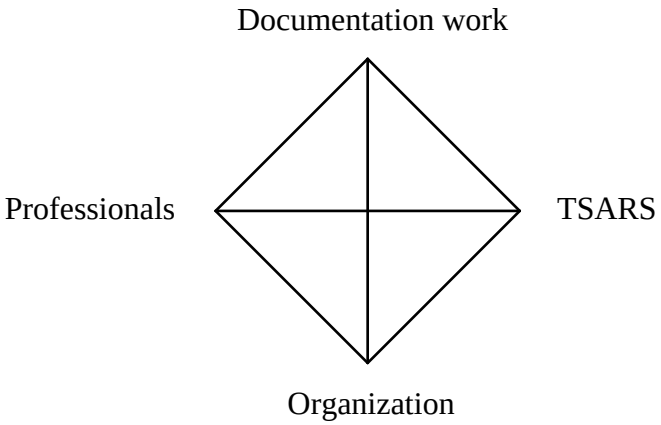


Figure 1.1. The four main elements in this study, each understandable only in its interplay with the others (adapted from Leavitt, 1964).

Two relationships in figure 1.1 are of special importance in this study. First, the one between the individual professional and the organization. The professionals and the organization have different perspectives on documentation work, and consequently their experience and assessment of the TSARS also differ. A major contribution of this study is to elaborate on this difference between a professional and a corporate perspective on documentation work and suggest certain implications for the development of TSARS. Second, the relationship between TSARS and their context. This relationship concerns the nature of computer systems in relation to human activity and, thus, defines the limits and contents of the term *computer support*. In describing computer systems as constructed models Naur (1988; 1990) stresses the principal distinction between constructions and beings, between models and the real world. Thus, while humans and computers might supplement each other in numerous ways they perform on fundamentally different premises, irrespective of whether or not these differences are made explicit in the design of the computer system. Handling these differences is a major challenge in providing computer support for documentation work.

### *1.2 Purpose and work done*

For three decades the efforts to provide computer support for documentation work have been primarily unsuccessful; the situation calls for a return to the origins to rethink the problem. This study attempts to accomplish this by attaching decisive importance to detailed, empirical investigations and by painstakingly separating observation and description from analysis and discussion. The entire study is divided into a number of independent studies with special emphasis on different aspects of documentation work and its computer support. These studies are documented in independent papers listed below, under the subpurpose to which they pertain. This summary is intended to set the independent studies into perspective and confront them with each other.

In its entirety, the purpose of my Ph.D. study has been threefold: First, to study practical, organizational contexts in which documentation work takes place, in search of its nature, peculiarities, and the current work practices. The objective is to understand—to listen—and, thus, a broad perspective is adopted, addressing the field of study in its full breadth. The work done concerns chemists doing research in a private enterprise and civil servants working in the central government:

- *Information Retrieval in a Work Setting: A Case Study of the Documentation Part of Chemists' Work* (Hertzum, 1993).
- *Computer Support for Journalization in the Danish Central Government* (Hertzum, 1994a).

Second, to emphasize change and diversity as persistent and unavoidable properties of professional work and suggest certain consequences for TSARS. Often, the need for changeability seems to be underestimated or sacrificed in favour of more immediate concerns, such as performance and development resources. The work done concerns the possibilities of involving the end-users in keeping TSARS up-to-date and the consequences of change on the choice of data model:

- *Information Retrieval Systems for Professionals: A Case Study of Computer Supported Legal Research* (Hertzum, Søes & Frøkjær, 1993).
- *A Comparison of Three Data Models for Text Storage and Retrieval Systems: The Relational Model Revisited* (Hertzum, 1994b).

Third, to evaluate the influence of the user interface on the quality and efficiency of documentation work. Interface design for TSARS requires both well-tested retrieval techniques with accompanying user interfaces and knowledge about how the users go about the interaction process. The work done investigates these issues experimentally in relation to browsing, querying, and a combination of these two retrieval techniques:

- *TeSS-projektet: Udvikling af et system til eksperimentel undersøgelse af brugergrænseflader til edb-baseret tekstsøgning* (Broløs, Frøkjær, Hertzum, Lárusdóttir, Pilgaard & Sørensen, 1993).
- *Browsing and Querying in Online Documentation: A Study of User Interfaces and the Interaction Process* (Hertzum & Frøkjær, 1994).

### *1.3 Outline*

Being a summary of the six papers mentioned above this report is entirely dependent upon the work reported in them; however, this report is intended to constitute a united whole which can be read independently. Five of the papers are articles and included here as appendixes; the sixth is a technical report available from DIKU. The reader may wish to read these papers first or consult them during the reading of this report.

The next section, section 2, deals briefly with the method. The forming and overall organization of my work is touched upon, and the methodological approach taken in connection with each of the subpurposes is characterized.

Section 3 through 5 constitutes the main part of this report. These three sections concern the main documentation activities—filing, managing, and retrieval. All three sections start by discussing the activity as it is perceived and organized in the interplay between the professionals and the organization, and all three sections end by discussing suggestions for supporting the activity by means of computers. Section 3 on filing deals with two different perspectives on documentation work, the professional and the corporate, and their implications for how filing is performed and why. Section 4 on retrieval deals with how retrieval is organized, how heterogeneous it is, and how it is affected by the user interface of TSARS. Section 5 on managing and modifying deals with the efforts performed to bridge or reduce the gap between the time and context in which a document is filed and the time and context of a searcher to whom the document is relevant.

Section 6 is the conclusion, and after that follows the references, a Danish abstract, and appendix A through E containing the five articles.

# 2. Method

Methodologically, the fundamental decision in my Ph.D. study has been to write articles rather than an all-inclusive thesis. I have found this approach attractive primarily for three reasons: It has created a number of structuring and motivating mile stones; it has provided opportunities for evaluation of parts of my work by competent people; and it has been open to doing work in cooperation with others. Furthermore, this approach has separated the overall design and planning of my Ph.D. study from the execution of the individual studies. This section deals with the overall design and planning; the methods used in conducting the individual studies are treated in the corresponding papers.

## *2.1 The purpose—a moving object*

When I started my Ph.D. study 2 1/2 years ago, I did not expect it to end where it is today. My understanding of the domain I initially set out to investigate has evolved significantly, and each step in this pursuit of the central issues has affected the purpose of my work. Furthermore, I have allowed opportunities to carry out sound studies of challenging issues considerable influence on the directions taken. Thus, the focal point of my work has been determined by an evolving understanding of the subject and by opportunities arising along the way. This has caused both minor and major modifications of the purpose. Two of the most thorough changes have been:

- I set out to study information retrieval, but gradually came to feel that this focal point missed much of the dynamics. This lead to the recognition of documentation work as the central task and, thereby, to understanding retrieval in its interplay with filing.
- In the beginning the main emphasis was on the literature and on discussing promising ideas theoretically; I mostly considered empirical studies as examples. However, I got stronger and stronger doubts about the speculative nature of this approach and, instead, adopted an approach driven by empirical and experimental work, though with room left for speculation.

## *2.2 The approach*

Empirical work does not escape being a human activity and, thus, inherently based on interpretation. However, by studying single cases it becomes a realistic objective to keep the interpretation loyal to the situation throughout the description and analysis. At other

times work gets more speculative, for example when interpretation is used actively to select and form changes which seem promising and worthy of further work. I have combined empirical and speculative work by carrying out three types of studies: work analysis, in connection with the first subpurpose of my Ph.D. study; idea exploration, in connection with the second subpurpose; and a controlled experiment, in connection with the third subpurpose.

*Work analysis*. The purpose of this activity is to study aspects of work as they are carried out in relevant, real-world settings. If successful, work analysis results in a coherent description with the, in a way, modest intention of giving an account of work as it takes place. Work analysis is closely related to both systems analysis, see for example Sommerville (1992), and task analysis, see for example Diaper (1989), but the term work analysis is chosen to distinguish it from both. Work analysis is opposed to systems analysis in that focus is on the work being done, rather than the systems involved or about to be developed. The difference is one of emphasis; while systems analysis is oriented towards systems development, work analysis is oriented towards the situation in which the systems are used. Systems analysis involve structured techniques such as entity-relationship modelling and dataflow diagramming, but also informal ones closer to my work analysis, e.g. interviews to acquire knowledge about how people perform their work. The outcome of my interviews is maintained as descriptions of concrete work routines and statements about problems, advantages, and personal attitudes. This distinguishes my work analysis from task analysis in which tasks are decomposed into subtasks and classified, as is task knowledge, in hierarchies and other taxonomies. Again the difference is one of emphasis; some of the approaches to task analysis are rather close to work analysis.

*Idea exploration*. The purpose of this activity is to work out an initial idea in significantly more detail to get indications of its costs, benefits, and viability. The idea is elaborated, it is not subjected to a formal test. Merely exploring an idea establishes it as a conception and, thus, affects our interpretation of the current situation. In this sense, a new situation is created. The strength of idea exploration is exactly the space it provides for working with changes of the current situation. This may be done by confronting the idea systematically with the literature or by producing a coherent description of the idea in a relevant context, for example by developing a system prototype.

*Controlled experiments*. The purpose of this activity is to test a hypothesis. The part of a controlled experiment directed towards formulating the hypothesis has much in common with idea exploration, but by performing an experiment it is made possible to accept or reject the hypothesis based on statistical evidence. This does not imply that the results are more well-founded than those of, say, work analysis; they are simply quantitative rather than qualitative. Quantitative methods require that everything involved

in the experiment is reduced to strictly defined, countable measures. Often, well-established conventions guide this quantification, but it is inherently an interpretation process. Therefore, careful qualitative interpretation of the results is needed in order to explain them, provide reasons to believe that the quantification is reasonable, and point to its limitations.

# 3. Filing

Filing is a varied activity ranging from capturing the thoughts of an individual to recording the mail entering and leaving an organization. Large individual differences exist, both between professionals and between organizations. These differences have important implications for TSARS and will be treated in section 5. This section concerns a difference which recurs in my studies despite the individual differences, a difference in the way filing and documentation work in its entirety is perceived by the professionals at the personal level and by management at the organizational level. Filing concerns units at all levels but these two stand out by holding different perspectives. In Hertzum (1993) the perspectives are referred to as the task-perspective and the document-perspective; to emphasize the actors holding the perspectives these terms are in the following replaced by *the professional perspective* and *the corporate perspective*, respectively.

These two perspectives differ in manifold ways and often interfere with each other. This results in filing being done for a combination of internal and external reasons. The professionals file to support their own working process and because they are told to; likewise the corporate perspective is motivated both by filing being useful to the organization and by external demands such as those laid down by the legislation. When filing is done for internal reasons it is experienced as meaningful and closely linked to the primary work; when it is done for external reasons it tends to be experienced as opposed to the primary work. In the following the nature and consequences of the interplay between the two perspectives are investigated, based on Hertzum (1993; 1994a).

## *3.1 The professional versus the corporate perspective*

The professional and the corporate perspective on the documentation work differ to the extent where they could almost be considered unrelated, provided the interplay between them could be neglected. While the professional documentation work exploits the properties of knowing the documents involved, the corporate documentation work enforces the prerequisites for searching without this knowledge. This distinction between a known universe and an unknown universe—the labels used in the following—is the key to understanding the perspectives and the conflicts arising from their coexistence.

The professionals know the documents they keep in their offices, hence retrieval of these documents means re-locating them. The documents filed are typically organized

according to their relation to the professionals' past and present work, for example documents pertaining to a certain case, product, or forthcoming meeting. Thus, retrieval is essentially turned into a process of recalling this relation. In this connection it is of minor importance that it might be necessary to leaf through the file or pile containing the document to actually get hold of it. The known-universe situation allows, and causes, the professionals to emphasize retrieval over filing. As retrieval is immediately valuable to the professionals it seems more worth the effort to spend five minutes retrieving a document, than to spend five minutes filing it. Consequently, filing is organized in a way that minimizes both the effort required to file individual documents and the maintenance of the system which determines where each document is placed.

The corporate perspective focuses on filing because, in an unknown universe, careful filing is the prerequisite for subsequent retrieval. The dependence on filing leads to instructions on how filing shall be done, to prespecified fields to be filled in for every document, and more and more often to the acquisition of computerized tools for supporting and institutionalizing the filing activities. Often, filing seems to become somewhat isolated, a tendency which is, for instance, reflected in the functionality of the journalization systems available in Denmark (Hertzum, 1994a). As another example, the filing instructions in Novo Nordisk require that a document is registered in the documentation system as soon as the document comes into existence (Hertzum, 1993). While this is natural in some situations and with some of the attributal information registered, it is usually premature to assign keywords to a document just created. However, from the corporate perspective biased and inadequate keywords are considered better than no keywords for two reasons: (1) They provide the major possibility for searching by subject, as opposed to by date, author, reference number etc. Subject retrieval can also be performed by full-text queries provided the documents are stored online, but as of yet this facility is only found in few organizations. (2) The keywords can always be modified later, an activity which is likely given the corporate perspective but significantly less so given the professional perspective. In the Emergency Management Agency, one civil servant cares little about the keywords attached to his cases in the journalization system, he prefers to maintain a personal register of his cases (Hertzum, 1994a).

The interplay between the professional and the corporate perspective is deeply affected by the nature of the primary work. In Novo Nordisk the documents in the chemists' offices dominate over those in the archive because the chemists keep a copy for easy reference when they file their documents in the archive. Thus, the archive contains mostly redundant data—in this respect the professional and the corporate perspective seem to duplicate, rather than supplement, each other. In, for instance, the Work Injury Office the civil servants seem to use the corporate filing routines more actively in the

organization of the documents relevant to their daily work (Hertzum, 1994a). Thus, in the Work Injury Office the two perspectives do not give rise to the same amount of redundancy. Presumably, the major reason for this difference is to be found in the nature of the primary work; while the chemists at Novo Nordisk do research the civil servants in the Work Injury Office decide cases. The chemists' work is the more free and creative; the civil servants' the more constrained and regulated. A superficial indication of this is that in Novo Nordisk documents are spoken of as the chemists' documents, while in the Work Injury Office they are referred to as the documents of the cases. It should be noted that the Work Injury Office has been established to handle certain, prespecified tasks; in other institutions the civil servants exercise more freedom in their work. Also, other chemists exercise less freedom than those studied in Hertzum (1993).

*3.2 Managerial incentives versus professional motivation*

Three incentives underpin the corporate perspective, a long term, a juridical, and a strategic. The long term incentive is to build a collective memory which provides access to the work of past and present employees. This incentive attacks the severe managerial problem that knowledge resides with the individual employees and requires their presence to be utilized. This problem can be reduced by carefully documenting the work done, e.g. by recording information about the actions taken, the results achieved, the persons involved and the like. From these data others can make their own interpretations, compare them to the original conclusions, and get in contact with persons who were involved provided they are still in the organization. Some people consider computers capable of storing knowledge and strive to develop knowledge elicitation schemes, for example to capture the employees' experience and expertise in corporate knowledge bases. However, the results these people have achieved can not justify their high hopes. Rather, knowledge seems to be an ability which becomes visible in action, not a thing which can be spelled out in documents, rules or the like, see for instance Dreyfus & Dreyfus (1986) and Naur (1988). Consequently, filing is a way to reduce, not solve, the problem of retaining acquired experience in the organization.

The juridical incentive differs substantially from public institutions to private enterprises. Public institutions are under an obligation (1) to take care that citizens on request get access to documents made or received by the institutions and (2) to periodically hand over their closed cases to the Danish National Archives to ensure that material from and about the public administration is preserved for future research. To live up to this obligation incoming mail, outgoing mail, and other documents pertaining to the cases must be carefully registered and filed. Novo Nordisk and other private enterprises need equally careful documentation when they want to apply for a patent or dispute

indictments of violating others' patents. In such cases it is essential to be able to provide evidence on when one first described or used a certain method, artefact or the like.

The strategic incentive has both an internal and an external part. Internally, new TSARS can be used as vehicles for organizational changes by enabling extraction of new information and by making information directly available to new persons. As access to information is often a source of power (Attewell & Rule, 1984), TSARS have the potential to bring about substantial changes. Externally, the strategic incentive consists in using the handling of the documentation work to promote the organization as serious and trustworthy. Some organizations display no strategic incentive and merely automate some previously manual procedures; others reorganize the work of the entire organization around the documentation tools, see for example Hertzum (1994a). Currently, strategic use of documentation efforts benefits from the popularity of quality standards such as ISO9000 and GLP (Good Laboratory Practice)—it is fashionable to induce some bureaucracy on professionals.

The motivation underlying the professional perspective is short term, personal, and pragmatic. The professionals' documentation work is woven into their primary work and mostly concerns ongoing projects and cases. The way they organize their desks visualizes the different tasks demanding their attention and, to some extent, constitutes their basis for answering the question fundamental to the organization of their daily work—what to do next. The tools used in this part of the documentation work, primarily piles arranged on the desk, are admirably simple and reliable to an extent where filing practically triggers retrieval. When a document is no longer needed on the desk to remind the professional of something, it is moved to the shelves. The material on the professional's shelves is organized to provide him with convenient access to it, neither to make it available to others nor to provide evidence of what have been done. When a folder looses importance it is removed from the shelves most easily accessed. In Hertzum (1993), the top shelf in one of the chemist's office contained a couple of folders left by the person previously occupying the office and a number of folders pertaining to the present chemist's past projects. Some documents remain in the office, for instance, because of general relevance or to witness about personal achievements, but it seems that the majority is eventually thrown away to make room for new documents.

### *3.3 The interplay*

Despite the fundamental differences between the professional and the corporate perspective they coexist. The typical way to organize filing seems to be (1) to leave it entirely to the professionals to organize their offices, i.e. corporate filing includes neither assumptions about how it is done nor advice on how it could be done and (2) to delegate corporate filing to secretaries or specialized units, i.e. the professionals' responsibilities

towards corporate filing is reduced to a minimum. This approach acknowledges the importance of both perspectives, but the differences between them are to a very large extent managed by refraining from exploiting the indisputable overlap between the perspectives. The price paid is a certain duplication of work and almost no exploitation of possibilities brought about by the other perspective.

With respect to the professionals, the key to explaining this lack of mutuality seems to be that they supplement the documentation of their own activities with personal contacts with other professionals, rather than with searches in the corporate archive. Their colleagues provide them with answers to specific questions, keep them aware of other projects and persons in the organization, remember previous efforts to achieve goals similar to those of a current project and so on. This is an informal kind of documentation work, and from the professional perspective it is often very effective. A principal reason for this is that in talking with a colleague both parties contribute to the interpretation of the problem. Searching in the archive it is the responsibility of the searcher alone to specify his problem, to select the relevant documents among those retrieved, and to understand these documents. While the documents stored in the archive are themselves valuable, it should be noted that, for example, in Novo Nordisk a major purpose served by the archive is to mediate personal contacts among the chemists (Hertzum, 1993). In this case the chemists capitalize on an opportunity which to some extent must be characterized as a side effect of corporate filing.

With respect to the organization, the lack of efforts to bridge the gap between the professional and the corporate perspective could be an attempt to gradually force the corporate perspective on the professionals. In part, this seems to be what has happened in the Work Injury Office where the organization of the primary work is tied very closely to the new journalization system which enforces the corporate perspective (Hertzum, 1994a). However, the approach taken in the Work Injury Office is exceptional, among other things because the primary work is much more rule-based and regulated than it is in, for instance, Novo Nordisk. In most cases the professional perspective is not under pressure from the corporate perspective, and the possibility of gaining mutual advantage from bridging the gap between the perspectives is either not recognized or not considered worth pursuing.

Contrary to this, the interplay between the two perspectives is in this study considered of strategic importance and, thus, central to success in providing computer support for documentation work. Emphasizing this interplay brings into focus the need for facilities, work routines and the like which contribute to bridge the gap between the perspectives. While the existence of the two very different perspectives on the documentation work is derived from empirical studies, the importance of facilities bridging the gap between them is only a hypothesis. However, it accords with Leavitt (1964) who identifies the interplay and interdependence between people and

organizational structure as highly important. Also, it is supported by having a well-balanced view on documentation work, balanced between short term and long term aspects and between a known-universe and an unknown-universe situation. Moreover, computers seem to provide substantially improved possibilities for obtaining mutual benefit from the efforts done in connection with each of the perspectives.

### *3.4 Computer support for filing*

Based on the above description of how filing takes place this section suggests two consequences for TSARS. First, a Janus-faced approach to the development of TSARS in order to bridge the gap between the professional and the corporate perspective. Second, background filing which concerns the possibility of automating a large part of the base registration of the documents.

#### *A Janus-faced approach*

It has long been recognized that individuals need support in organizing their documents and other information. Bush (1945) presents one of the first and most influential visions of such a personal information system. Many systems have continued along the lines of Bush's Memex, including seminal systems such as Augment (Engelbart & English, 1968) and Dynabook (Kay & Goldberg, 1977). In particular, the proliferation of personal computers during the 1980s has been accompanied by a vastly increased interest in and demand for information systems for personal use, see for example Mander et al. (1992). These systems share the assumption that the user is an independent actor, i.e. is not embedded in an organizational context. Moreover, they mostly address electronic documents only and, thus, require either that all documents are available in electronic form, for instance through scanning or over a computer network, or that non-electronic documents are handled separately.

It is equally well-recognized that organizations need systems—manual or computer-based—to ensure precision and care in their document handling. However, corporate filing seems to be considered too trivial to be treated in its own right in the literature. Usually, corporate filing is considered a minor aspect of management information systems and focus is, instead, on the information which can be derived from the filed data and on management's utilization of this information. Thus, the perspective is corporate or managerial and tends to be so throughout. Some studies discuss practical problems involved in organizing and carrying through the filing activities, but I know of no studies which discuss the corporate perspective in opposition to a coherent understanding of a contrasting, professional perspective on the documentation work.

In the literature focus seems to be on either the professional or the corporate perspective, and in practice these two perspectives seem to lead rather separate lives too.

However, this study hypothesizes that it is of strategic importance to bridge the gap between the perspectives. It is suggested that this is done by designing TSARS as two separate but interfaced systems, a suggestion which is a consequence of the profoundly different nature of the two perspectives. The interface specifies the information which must be made available by the professional for up-loading to the corporate part of the system and the information which can be down-loaded from the corporate part to individual professionals. This design maintains a boundary between the professionals' own documentation work and their part of the corporate documentation work, a property with at least two advantages: (1) By making this boundary visible to the professionals they will know when they use their system for their own benefit and when they fulfill their obligation towards the organization. (2) As it is singled out, the professionals' part of the corporate documentation work, identified by the information which must be up-loaded, can be minimized during system design.

The primary purpose of the information which is up-loaded is to provide a base registration, i.e. to make documents known to the corporate system. This base registration can be followed up by requests for supplementary information where necessary and by reminders if nothing further happens. However, a large part of the base registration need not involve the professionals at all; it can be performed automatically in the background, a solution with obvious advantages.

*Background filing*

From the corporate perspective the only way to ensure that the relevant documents are filed is to demand that everything is filed. In many places this has been achieved for incoming and outgoing mail as well as for major internal documents, but for many internal documents it is still an objective. Some documents are so informal that they are considered personal, for instance a list with the current fines for a frequent, standard type of legal case (Hertzum et al., 1993); others are so numerous that filing appears to be an overwhelming task, for instance the data sheets from the chemists' experiments at Novo Nordisk (Hertzum, 1993); and still others remain unfiled for various practical and habitual reasons, for instance e-mail messages. From the professional perspective filing everything is a time-consuming and for many documents pointless task. To generate and maintain the care necessary in filing the obviously more important documents it is essential not to flood the professionals with regulations concerning all the less important ones. One approach which takes this into account is delegation, e.g. filing is handled by secretaries or specialized units. However, this still requires that the professionals see to that all the documents they produce or receive reach a secretary or the specialized unit.

What appears to be needed is a quick-and-dirty or discount way of ensuring that all documents are filed combined with careful filing of selected documents. Even though

many internal documents exist in hard copy, a large part of them have been prepared by computer. Thus, catching all electronic documents would provide a rather good basis for the discount filing. This goal can be achieved automatically with programs similar to current backup systems which over a network copies files from the employees' personal computers to a central backup store. In addition to the document contents certain attributal information can be obtained: Author and date are readily available, the filename can be used as document title, and well worked-out techniques exist for the automatic selection of keywords, see e.g. Salton (1971; 1986) and Salton & Buckley (1988).

A background filing facility like this would provide an automatically updated working copy archive containing the full text of the documents. If such a facility appears to store masses of irrelevant material in addition to the important documents, management should consider carefully: (1) Whether this feeling stems from lack of confidence in the automatic data collection and indexing. In this case it should be emphasized that human intervention is necessary to select the important documents for special treatment and to file the documents not available in electronic form. (2) Whether this feeling has its roots in the demand that everything is filed. In this case the incentives for filing should be reconsidered, and in this connection the backup function inherent in background filing could become a new, major argument for all-inclusive filing.

Documents important to the organization can often be identified by type—technical reports are important whereas minutes from project group meetings are mostly not. Thus, one way to select documents for special treatment, i.e. human indexing, would be to require human indexing of certain types of documents. To capture additional important documents a straightforward possibility is to make the authors responsible for identifying them. In practice, the professionals could periodically receive a list with all the documents they had created during the period. For each document in this list the professionals should indicate whether or not it is important and, if important, it should pass through human indexing. Standard document setups which are invoked when a document is created to format it according to its type provide a way to ensure that the type of the documents is known to the system. Thus, the use of standard document setups would make it possible to automatically indicate in the list which documents must go through human indexing. Otherwise, it would be the responsibility of the professionals to identify these documents as well as any additional important ones.

# 4. Retrieval

Retrieval is the activity through which the recorded information is utilized. However, retrieval may also be the activity in which relevant or even pertinent information is overlooked, because it is not retrieved at all or because it is buried in retrieved, but irrelevant, material. To successfully retrieve a document the searcher must be able to formulate to himself what he is looking for and to express this request in a way which takes account of the differences between himself in his present situation and the past situation in which somebody wrote and filed the document. In the professionals' offices the past situation is reasonably well-known because the person requesting a document is the same as the one who once filed it. Retrieval in an unknown universe is substantially more difficult because little or nothing is known beforehand about the contents of the filed documents and the contexts in which they were produced. These difficulties can be reduced somewhat by carefully and continuously modifying the documentation system to bridge the gap between different professionals and across time, see section 5. However, change and diversity are conditions to which retrieval is inescapably subjected.

From the corporate perspective retrieval can serve strategic purposes, for example by measuring performance, or juridical purposes, for example when applying for a patent, but the major day-to-day purpose is to service the professionals. The professionals engage in retrieval to further their primary work, and this obviously agrees well with the major corporate purpose. Thus, the organization of corporate retrieval reflects an agreed upon objective but takes different forms depending on, for instance, the importance attached to retrieval and the nature of the primary work. Searches, whether in the corporate archive or the professionals' offices, are performed to achieve manifold ends, and these different types of searches are not affected uniformly by change. The professionals seem to take considerable advantage of this by organizing their documentation work in a way which enables them to replace some of the change-sensitive searches with less sensitive ones.

### *4.1 The organization of corporate retrieval*

TSARS have made it possible to separate and distribute the activities involved in the documentation work. Queries can be posed from any terminal or personal computer in the network, irrespective of where the filing activities are performed, where the documents are physically located and so on. The potential effect on the organization of retrieval is

very large; however, several other factors also affect the situation. One of these is the importance management attaches to retrieval. In many central government institutions the external demands laid down by the legislation seem to be a major reason for the overall organization of the journalization work; less attention has been paid to the technological possibilities and the civil servants' needs (Hertzum, 1994a). It should however be noted that the daily retrieval activities are directed specifically towards the civil servants.

Retrieval is almost always considered an activity initiated by and relevant to individual professionals. Thus, retrieval takes place on request from the professionals, whether they search themselves or delegate the task to secretaries or specialized units. If the professionals make no requests, no retrieval takes place. The exception from this rule is centrally driven retrieval where the professionals are fed with documents retrieved by so-called documentalists. Retrieval from the corporate archive and external sources are found to be organized in three ways—centrally driven retrieval, end-user retrieval, and delegated retrieval.

*Centrally driven retrieval*. A rare example of centrally driven retrieval is found in another part of Novo Nordisk than the one studied in Hertzum (1993). Here a number of documentalists are responsible for scanning books, journal issues, proceedings and the like to spot documents of interest to individual chemists. The documentalists take care of the searches intended to keep the chemists abreast of the latest developments within their respective areas of interest. Due to their information retrieval knowledge and daily practice, the documentalists can perform sophisticated searches which fully utilize the capabilities of the different retrieval systems, access the relevant ones and so forth. Furthermore, the documentalists go through the lists of documents returned from the retrieval sessions to pick the documents of interest and discard the rest. Thus, the chemists are relieved from scanning journals and the like to maintain awareness, and they are kept up to date on a basis which is comprised of more journals and other sources than they would monitor themselves.

This organization of retrieval builds on the assumption that the documentalists are able to recognize the documents of interest to the chemists. To do this successfully requires basic chemical knowledge and close acquaintance with the projects in which the individual chemists are involved. Each documentalist can span only a limited number of chemists, and care must be taken to ensure regular contact between documentalists and chemists. One obvious occasion for these contacts is the searches which the chemists initiate to answer specific questions, but unless the chemists work on long, stable projects additional contacts are of decisive importance. Moreover, this way of organizing retrieval is probably only cost-effective in large organizations which are very dependent upon being abreast with the latest achievements in their field.

*End-user retrieval*. Normally management does not turn retrieval into a corporate task. Rather, the professionals are often expected to be motivated to search themselves provided that tools to do it are at hand. The justification for this expectation is the immediate value attached to retrieval—contrary to filing retrieval is an end in itself. However, professionals in Novo Nordisk (Hertzum, 1993), in the Danish central government (Hertzum, 1994a), and in general (Mischo & Lee, 1987) tend to avoid end-user retrieval from corporate archives or external sources. One way to do this is to build personal archives and, thereby, convert unknown-universe searching into known-universe searching, a context in which end-user retrieval is something completely different and perfectly natural. However, personal archives are usually considered harmful from the corporate perspective because they tend to exist at the expense of, rather than in duplication of, corporate archives.

At present, the adoption of end-user retrieval is limited, though it has been introduced in some organizations and some professionals find it straightforward and useful. Four reasons for avoiding end-user retrieval have appeared in my studies (additional reasons can, for instance, be found in Mischo & Lee, 1987):

- Systems overload. TSARS are introduced to handle information and reduce so-called information overload, but the TSARS themselves constitute an increase in the amount of information to be processed and may thus cause a variant of the original problem, systems overload. BORIS, the new documentation system in Novo Nordisk, is an example. The problem is not one of user-friendliness; BORIS is rather straightforward. Nor is the problem caused by lack of experience with computers. The chemists do basic text processing almost daily, and some also use spreadsheets, special programs to do chemical computations etc. The problem is the ever-increasing number of systems available to the chemists: It is inconceivable to get acquainted with all these systems and hard just to find out which are worth approaching (Hertzum, 1993).
- Critical mass. When TSARS are introduced filing can immediately be moved to the new system, but it is usually impracticable to transfer the documents already in the old system to the new. Furthermore, filing is often not moved all at once. As an example, Novo Nordisk is moving from a hard copy archive to an electronic one but is explicitly not striving for storing everything in electronic form—they want an archive which is true hypermedia. In the beginning the old system contains all the documents whereas the new system contains none, and during a shorter or longer transitional period retrieval involves both searching in the old system and in the new. Before the new system is considered the major one it must reach a certain size, its critical mass. Until that happens many will hesitate to use the system, and as it is

usually the professionals' own decision whether they will use it or not end-user retrieval remains an infrequent activity.

- Manual retrieval often outperforms online retrieval. The term end-user retrieval usually implies retrieval by computer, and several comparisons of manual and online retrieval have turned out in favour of manual retrieval, see for example Marchionini (1989) and Hertzum & Frøkjær (1994). These experiments picture online retrieval as a demanding task which many users experience difficulties in dealing with. Forcing the users to search online has severe implications over allowing them to search manually. Thus, in the present situation, i.e. with contemporary TSARS and professionals without much experience in using them, reluctance towards end-user retrieval might be cost-effective.
- A lengthy process. Many efforts to provide computer support for documentation work have focused narrowly on automating previously manual procedures in which end-user retrieval was not an option. This illustrates how difficult it is to transcend the existing situation and rethink it given new possibilities provided by, for instance, technology (Hertzum, 1994a). A further reason for this tendency is that many professionals have only limited experience with computers and therefore hesitate to take on computerized tasks. Thus, the possibility of end-user retrieval has been slow to be realized.

*Delegated retrieval*. Instead of searching themselves the professionals prefer to delegate retrieval, i.e. to have it done by persons doing it sufficiently often to become and stay proficient. Delegated retrieval can be performed by secretaries or specialized units. In the Danish central government most institutions have specialized units which take care of the journalization work, including retrieval. The acquisition of journalization systems has not changed this situation much, though taken alone the small institutions exhibit a quite different picture in which the persons in charge of journalization are almost exclusively secretaries (Hertzum, 1994a). The facilities which make end-user retrieval an option for the professionals also provide the secretaries with the possibility to add a new, important task to their duties in place of for example typing. From a managerial perspective moving retrieval from a specialized unit to the secretaries might be considered a more flexible way to organize work.

### *4.2 Change and types of searches*

As already noted, change is an unavoidable property of professionals' work and a condition to which retrieval is inescapably subjected. However, there are different types of searches and they are affected to different extents. First, one must distinguish between known-universe and unknown-universe searching. Second, the objective of the search affects the influence of change. A simple and informative categorization of searches based

on their objective has been suggested by Meadow (1992); the four general types of searches he identifies are used in the following:

*Known-item searching* where the searcher knows exactly what documents are wanted and can specify them precisely by means of searchable attributes. Examples: "The case file with reference number 5110-29" and "All papers authored or coauthored by Ms. X who arrives here next week as a guest researcher". In such situations change is dealt with, consciously or unconsciously, prior to retrieval and presents no problem during retrieval. This robustness towards change is a major advantage of known-item searching and makes this type of searches significantly more straightforward than the others. One of the attractive properties of searching in a known universe is the high ratio of known-item searches, whereas this ratio is much lower in the case of unknown-universe searching.

*Fact retrieval* where the searcher is looking for specific information, but without necessarily knowing where to look for it. Example: "What was the size of the Danish *moms* (*moms* translates to VAT, the abbreviation for value added tax) when it was initially introduced?" It is not certain what terms to use for searching, but some initial candidates are readily available. In the example *moms* appears to be a natural choice of query term, but it does not occur in the legal texts defining the concept and its legal implications because it was considered colloquial at the time these legal texts were written (Hertzum et al., 1993). Thus, even though the searcher knows exactly what piece of information he is looking for, he may experience severe trouble in expressing his request with due regard to the intervening changes. Obviously, the searcher is better equipped to compensate for changes during known-universe searching than unknown-universe searching.

*Subject retrieval* where the searcher is looking for information on a subject in general. Example: "What is the state of the art on TSARS?" There is no one way to describe the subject, no one way the desired information will be represented, no reason to expect that it will all be contained in a single document, and virtually every reason to expect that some relevant information stay unrecognized. In the literature, the major retrieval technique for subject retrieval has traditionally been querying, but coping with change and diversity introduces severe problems, see e.g. Blair & Maron (1985). In a known universe, subject retrieval can often be replaced by a number of known-item searches and this way the problems with change are reduced to a manageable size. In unknown universes a step in the same direction is to perform subject retrieval by browsing rather than querying. Empirical studies of professionals' retrieval activities (Ellis, 1989; Ellis et al., 1993) indicate that this is common, and in the TeSS experiment (Hertzum & Frøkjær, 1994) it is found to be more efficient too.

*Exploratory retrieval* where the searcher intends to find out what kinds of information are available, not to answer a specific question. Example: On entering a library for the first time one often scans the bookshelves, not to see if a particular book is

present, but to see what kinds of books are found there. Exploratory retrieval takes place only in connection with unknown-universe searching and is the kind of retrieval most widely connected with the family of retrieval techniques known as browsing. Exploring is to some extent an activity that reveals changes, rather than suffers from them. However, for a searcher to become aware of a certain shelf in a library or engage in exploring a certain branch in a hierarchy of online documents, he must experience some connection between the appearance of the documents and his own interests.

In the literature subject retrieval in unknown universes is by far the most studied type of retrieval. Empirical studies such as Ellis (1989), Ellis et al. (1993), and Hertzum (1993, 1994a) suggest that subject retrieval is not nearly as dominating in practice. Two circumstances may help explain this: First, the nature of the primary work. In the central government, the documents needed to perform the primary work are usually those of the case in question, neither documents from other cases nor from external sources. This is reflected in the way filing takes place, e.g. documents are assembled in case files, and journalization systems have two levels, namely a case level with general information about the cases and a subordinate document level where the actual documents are registered (Hertzum, 1994a). To a large extent cases and their documents are retrieved by reference number, a matter of known-item retrieval. Searches which cut across case boundaries are perfectly possible, but they are not the rule. In Novo Nordisk the dependence on subject retrieval, especially from external sources, is considerably larger. The chemists do research in a commercial environment and need to be aware of the accomplishments of independent and competing researchers. This situation is much closer to the one which has inspired most of the work reported in the literature, i.e. most information retrieval researchers have taken their own situation as researchers as their point of departure.

Second, subject retrieval is often converted into known-item retrieval. One widespread way to achieve this is through the documents in the professionals' offices. These documents reduce the need for retrieval from corporate and external sources, and—as already mentioned—subject retrieval in a known universe often amounts to little more than a sequence of known-item searches. Another important way to avoid subject retrieval is to utilize personal contacts. Rather than start searching from scratch the professionals turn to colleagues for advice on which documents to read, which authors to pay special attention to, which sources to check and so on. If a colleague knowledgable on the particular subject is found, this is a highly effective kind of retrieval which can often add a lot of context, concrete suggestions, and clues to the primary work as well as to the searches.

*4.3 TeSS—computer support for retrieval*

Karnov's Law Database (Hertzum et al., 1993) provides facilities for known-item searching as well as subject retrieval. Especially, it includes a facility designed specifically from the idea of turning subject retrieval into known-item searching, namely the so-called reversed indexing facility where terms in the thesaurus are used as keywords with visible references to legal texts of special interest. TeSS (Broløs et al., 1993; Hertzum & Frøkjær, 1994), the system in focus in this section, is directed towards subject retrieval.

In section 3 it was assumed that background filing had to be supplemented with human indexing of the important documents to ensure the necessary quality. This assumption was based on philosophical arguments about the nature of knowledge. The TeSS experiment provides an opportunity to compare these arguments with experimental findings. Also, in Hertzum & Frøkjær (1994) the experimental data are analyzed with respect to differences between the retrieval modes involved in the experiment. However, the differences between the modes might hide large individual differences between the users, for example some users may be consistently better at querying than browsing while others are examples to the contrary. This section investigates these two aspects of computer support for retrieval—full-text versus human indexing and individual differences.

*Full-text versus human indexing*

From the point of view of comparing automatically prepared and manually prepared retrieval the TeSS experiment includes a comparison of full-text retrieval and retrieval based on a kind of human indexing. Two retrieval modes, LOGICAL and VENN, provide full-text retrieval, i.e. retrieval where queries are matched against the words occurring in the document contents. Thus a document may be retrieved on the basis of any word appearing in it, with the exception of a small number of very common words, such as *a*, *to*, and *of*. Human indexing consists in the manual selection of a small number of words which characterize each document. Two retrieval modes provide retrieval based on such keywords: First, a hard copy version of the manuals, PAPER, provides access to the texts through the table of contents and the index which are comprised of words selected by humans to either characterize the contents of the sections or pinpoint where specific topics are treated. Second, a browsing mode, BROWSE, provides access to the table of contents through facilities to both move up and down the hierarchy of headings and search it for specific words. Finally, a retrieval mode combining browsing and querying, ALL, provides access to the table of contents and offers full-text queries.

The experimental results agree well with the philosophical arguments holding that human indexing provides improved possibilities for successful retrieval. PAPER displays

the lowest average task completion time and the solutions of highest average quality, and among the four computer-based retrieval modes the lowest average task completion time is achieved with BROWSE while no significant differences exist with respect to quality of solutions. It should be noted that while BROWSE seems to capitalize on a kind of human indexing this is done with no extra human effort. Most technical writing has informative headings, and once the text is written they can be exploited automatically for retrieval purposes.

The information retrieval literature contains many experimental comparisons of techniques which attempt to support retrieval by exploiting various document properties. The results of these comparisons are rather inconsistent, but one reasonably general finding is that full-text retrieval is to be preferred when most relevant documents should be retrieved whereas keyword-based retrieval is preferable when a few, central documents are needed, see for example Tenopir (1984), Lancaster et al. (1989), and Tenopir & Ro (1990). In the TeSS experiment the solutions of the tasks are closer to requiring a few, central texts than the linking together of information from a broad range of texts. Thus, the TeSS experiment accords with the above finding. However, this finding also implies that a more complete retrieval system would result from combining full-text retrieval and keyword-based retrieval. In the TeSS experiment this combination was represented by ALL and the performance costs of its completeness were significantly larger than the benefits. The users in the experiment seemed to suffer from an overload very similar to the systems overload encountered as a reason for avoiding end-user retrieval.

*Individual differences*

Standard statistics to characterize individual differences consist of measures to express the span between an upper-end and a lower-end performer with respect to, for example, task completion time. Two much used and readily interpretable measures are (Egan, 1988): The *maximum/minimum ratio* which measures the relationship between the two extreme users in the experiment. However, it is a rather unstable measure because it is based on only two, especially variable data points. A more stable measure is the *75th percentile/25th percentile ratio* (the 25th percentile is the value for which 25% of the users have better performance and 75% have worse, and the 75th percentile is the value for which 75% of the users have better performance and 25% have worse). Egan (1988) reviews a number of studies in which task completion time was measured and finds that for information retrieval the maximum/minimum ratio is approximately 9:1 which is more than for text editing and less than for programming. For all three kinds of task the 75th percentile/25th percentile ratio is close to 2:1.

In the TeSS experiment these statistics can be used both to assess the magnitude of the individual differences and to investigate whether they vary from one the retrieval

mode to another. Each user's performance with each retrieval mode is the average of the four tasks solved with that mode. The individual differences are computed for task completion time, and the results, in terms of the maximum/minimum ratio, the 95th percentile/fifth percentile ratio, and the 75th percentile/25th percentile ratio, appear in figure 4.1. Without attaching too much weight to the maximum/minimum ratio it should be noted that it varies substantially with the retrieval mode. The modes obtaining the best overall performance are those for which the span of the individual differences is small. The 95th percentile/fifth percentile ratio is included here as an alternative to the maximum/minimum ratio. Like the maximum/minimum ratio, the 95th percentile/fifth percentile ratio is a measure of the difference between extreme users, but it is statistically more reliable. For LOGICAL, VENN, and ALL a comparison of the maximum/minimum ratio and the 95th percentile/fifth percentile ratio reveals large performance differences among the 5% of the users having the best performance and among the 5% having the worst performance. Also, the 95th percentile/fifth percentile ratio bears witness to considerable individual differences, and they are consistently larger with the modes of TeSS than with PAPER, the retrieval mode drawing the most from the users' thorough familiarity with books. The 75th percentile/25th percentile ratio is the same for all retrieval modes and equal to the one found by Egan (1988). Broadly speaking, an average upper-end performer solves a task in half the time an average lower-end performer needs. However, the costs or benefits that might result from assigning different persons to a task are much greater, and there is reason to believe that they will be even greater in practice than in the TeSS experiment where the users were much alike in terms of age, education and the like.

| | max/min | p95/p05 | p75/p25 |
|---|---|---|---|
| BROWSE | 11:1 | 6:1 | 2:1 |
| LOGICAL | 28:1 | 8:1 | 2:1 |
| VENN | 24:1 | 6:1 | 2:1 |
| ALL | 39:1 | 7:1 | 2:1 |
| PAPER | 6:1 | 4:1 | 2:1 |

Figure 4.1. Individual differences in task completion times.

The above measures of individual differences determine their magnitude but make no attempt to explain them. One possible explanation is that the differences between the retrieval modes match the abilities of different users: Maybe some users are consistently better at manual than computer-based searching while others are examples to the contrary; maybe some users are consistently better at querying than browsing while others are examples to the contrary and so on.

To test whether this is the case each user's performance with each retrieval mode was characterized by being assigned to one of three performance groups. The first group contained the 25% of the users having the best performance, the third group the 25% having the worst performance, and the second group the remaining 50%. From these data it is possible to make pairwise comparisons of the retrieval modes to see, for instance, whether the 25% of the users performing best tend to be the same for both modes.

The pairwise comparisons of BROWSE and LOGICAL and of BROWSE and VENN are given in figure 4.2. It can be concluded that browsing abilities and querying abilities do not differentiate among the users, i.e. very few users are in the first performance group for one retrieval mode and, at the same time, in the third for the other mode. On the contrary, users good at browsing tend to be good at querying too, and likewise users bad at browsing tend to be bad at querying too. This pattern recurs for all the other pairwise comparisons. Thus, though the individual differences are to a certain extent affected by differences between the retrieval modes, they are primarily determined by other factors. Experience, domain knowledge, and technical aptitudes have proved important factors in other studies (Egan, 1988; Borgman, 1989), but individual differences is still a relatively poorly understood area.

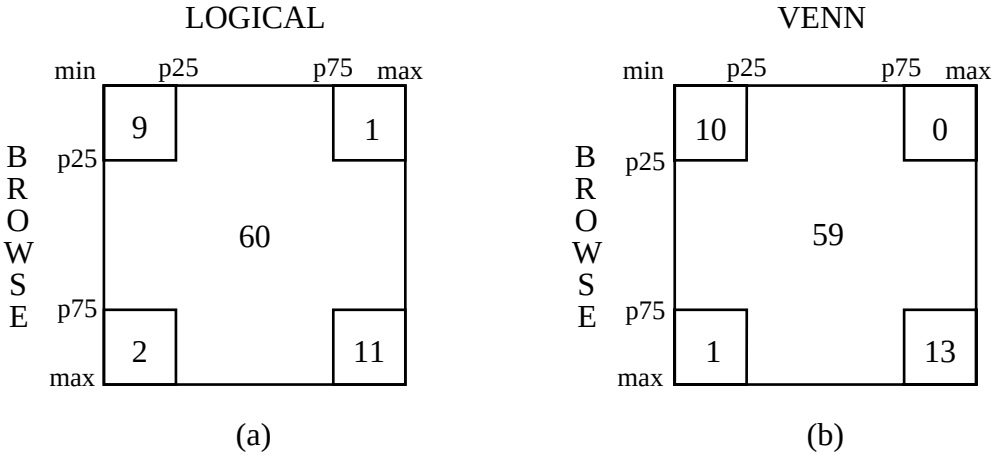


Figure 4.2. Browsing performance correlated with querying performance, in terms of the number of users with the different performance combinations. The pictures tend to be symmetrical around the diagonal, i.e. browsing and querying performance tend to go together. (a) BROWSE - LOGICAL; (b) BROWSE - VENN.

# 5. Managing and modifying

Once established, tools and work routines to handle documentation work, or any other task, tend to define it too (Naur, 1965). As long as the tools and work routines are appropriate, filing and retrieval can take place according to them with no need for additional activities to manage the documentation work. However, when the documentation work needs to be adapted to new circumstances it takes explicit action to modify the tools and work routines. The process which consists of monitoring work, technological developments and the like to decide when modifications should take place and of carrying through these modifications constitute the management of the documentation work.

One incentive to perform modifications is dissatisfaction with the current organization of the documentation work; another is visions about how, for example, new tools can be exploited to the advantage of the documentation work. As far as the acquisition of journalization systems in the Danish central government is a typical example these incentives are only rarely present simultaneously (Hertzum, 1994a). A certain dissatisfaction combined with very limited visions is much more common and leads, to a considerable extent, to uncertainty and hesitation rather than modifications. In the literature discussions of the management of tools and work routines for documentation work tend, like the literature on filing, to assume either the professional or the corporate perspective. Thus, either the individual user is allowed to make any modification he likes, the approach adopted in most hypertext systems (Halasz, 1988), or all modifications are under centralized control, the predominant approach in the information retrieval community (Hertzum et al., 1993). In general, the management of the documentation work is surrounded by much uncertainty and there are few well-documented examples to learn from.

## *5.1 The need for modifications*

Change and diversity are persistent and unavoidable properties of professionals' work. Mintzberg (1983) emphasizes the ability to perform in changing and complex environments as a distinguishing characteristic of professionals. If the environment is neither changing nor complex there are more cost effective ways of organizing production than relying on professionals. When the environment changes it also affects the

professionals' documentation work. On the one hand the professionals try to adapt their documentation work to the new situation; on the other hand TSARS and other documentation tools reflect the old situation and tend to enforce status quo. As a result, the documentation work gets somewhat removed from the primary work. Filing becomes more tiresome and less precise because the structure and assumptions built into the documentation system match the documents only partially. Retrieval becomes harder because filing is less precise and because it is left to the searcher to take account of the changes that separate his present situation from the past situation in which the wanted documents were filed.

The diversity which arises from differences between the individual professionals makes it impracticable for specialized units or other organization-wide bodies to keep TSARS in a state where they match the individual professionals' primary work. Furthermore, efforts in this direction would result in pronounced diversity and, therefore, be in opposition to the corporate perspective which strives for standardization in order to establish a common structure facilitating retrieval. From the corporate perspective an important consideration is to incorporate changes in a way which, to the greatest extent possible, is transparent during retrieval. This means, for instance, that day-to-day modifications of a keyword list are avoided because they complicate the choice of query terms. Changes are usually collected over a period of time and then introduced collectively in terms of a new version. From the professional perspective the major objective is to keep the way the filed documents are organized in accordance with the primary work and this is practically always done through day-to-day modifications rather than a sequence of distinct, internally consistent versions. Two central reasons for this are that the consistency achieved with versions is not decisive in a known universe and that the overhead involved in producing the versions is considered too big.

In some situations the primary value of, for example, a thesaurus is that all users have identical copies of it. An obvious example being that indexers and searchers should consult the same thesaurus when assigning keywords and selecting query terms, respectively. In these situations the thesaurus provides a common point of reference easing the coordination of cooperative work, and it is crucial that modifications are under centralized control. In other situations the value of a thesaurus is dependent upon the possibilities of making it mirror the user's specific and continuously changing situation. Here the thesaurus is perceived as a personal or project group tool, and facilities to perform local modifications are essential. Mostly, it would be desirable if the thesaurus was, in some way, both common to many users and open to local adjustments. This would (1) provide individual professionals and project groups with a basic thesaurus from which to grow their own special-purpose versions and (2) cast and store these local thesaurus versions in a way known by many and available to the organization. This issue is taken up

again below, after a discussion of the viability of allowing the end-users to perform modifications themselves.

*5.2 End-user modifiability*

In the past, most computer systems were based on centralized architectures segregated from the persons who needed to apply them in their daily work. In continuation of this the systems were perceived from the corporate perspective and their evolution was necessarily organized as a process under centralized control. With the proliferation of personal computers during the 1980s a large number of people have become involved in other computing tasks than mere usage of ready-made systems, a phenomenon known as end-user computing, for reviews see e.g. Brancheau & Brown (1993) and McLean et al. (1993). One of several approaches to end-user computing is end-user modifiability. Though the term emphasizes the individual professional, facilities for end-user modifiability might just as well be used by cooperating groups of professionals or through delegation.

End-user modifiability means providing individual users or groups of users with facilities which enable them to adapt a system to fit their needs and preferences. An example of an end-user modifiable system is Karnov's Law Database with its dynamic thesaurus, dynamic classification structure, and personal notes (Hertzum et al., 1993). The appropriateness and success of end-user modifiability depend on (1) the users' ability and inclination to devise appropriate, local modifications; (2) the match between the modifications the users want to make and those they are enabled to make; and (3) the reconciliation process responsible for integrating the modifications performed by the end-user and those introduced by new versions.

Critics of end-user modifiability focus on the first precondition, i.e. they doubt end-users will be able to perform modifications without introducing many inconsistencies and errors, see for example Tague (1981), Mason (1986), and Tyler & Treu (1989). Admittedly, the users are in most cases required to be professionals and thereby subject specialists. Above that I am convinced that professionals can make modifications which are sufficiently relevant and correct to make end-user modifiability valuable. On the other hand, the experiences with end-user retrieval suggest that professionals will not care to spend time modifying TSARS—in many cases professionals are reluctant to use TSARS at all. It seems likely that some professionals will make personal modifications, but indirectly by instructing secretaries to do it; and in a project group local modifications of a thesaurus could be an important contribution to the establishment of a group language. The value of end-user modifiability is not determined by whether or not the professionals operate the facility themselves, rather it should be remembered that without end-user modifiability the professionals are forced to either discard the system or adapt to it.

The second precondition seems to me to be more problematic. Fischer & Girgensohn (1990) state that: *"End-user modifiability is not a luxury, but a necessity in cases where the systems do not fit a particular task, a particular style of working or a personal sense of aesthetics."* However, at the time TSARS are constructed it is impossible to foresee which modifications future users will want to perform. Thus, there is no way to secure that the assumptions forming the unchangeable backbone of the system are not in conflict with some user's task, style, or aesthetics. Potential topics for end-user modifiability include but are not limited to:

- modifications of a thesaurus or keyword list,
- modifications of a classification structure,
- annotations adding various comments to the documents,
- links establishing connections between documents,
- new standard document setups, and
- definition of rules enforcing certain integrity constraints.

As more experience is gained with facilities for end-user modifiability, the kinds of modifications which can be foreseen and supported will probably become the common ones relevant to many users. At present empirical studies are few, one is reported in Jørgensen & Sauer (1990).

The third precondition for the success of end-user modifiability is that personal modifications must not preclude the use of any feature that might have been used provided the user had not made the modifications. This means, among other things, that the user must be able to upgrade the system to a new, centrally prepared version without loosing his personal modifications, otherwise he is forced to decide whether he prefers the new version or his personal modifications. Thus, an essential aspect of end-user modifiability is the integration of local modifications and new versions. This problem is also known from hypertext systems (Halasz, 1988) and from manufacturers' maintenance of application programs existing in several versions tailored to individual customers and modified both to meet individual customers' needs and to introduce new general facilities. Fully automatic solutions to the problem do not exist because a significant part of it consists, not in deciding how to combine two modifications, but in deciding which one of two conflicting modifications the user prefers. These conflicts pertain, in many cases, to differences between the professional and the corporate perspective, for example conflicts between immediate needs and general relevance. The problem deserves much more attention.

### *5.3 Computer support for managing and modifying*

Presumably, the most studied aspect of the management of documentation work is thesaurus management. While the conventional approach to this activity is manual, new

possibilities are emerging as more and more thesauri are included in TSARS. However, the practical impact of computer support for thesaurus management has, so far, been very limited. The second topic treated in this section concerns the tools used to build TSARS. These tools should lend themselves to the development of TSARS which are capable of being modified. Two ways to strive for this are discussed and contrasted, generic TSARS and general-purpose toolkits.

*Thesaurus management*

Some TSARS include a thesaurus as a terminological aid between the user and the documents. A thesaurus can be thought of as a map of the terminology in a given field (Foskett, 1985). In essence, it is a network of terms connected with links. A thesaurus often constitutes a hierarchy made up of *broader than* and *narrower than* links; further, *related to* links connecting synonymous or nearly synonymous terms are common, as is notes giving information about e.g. term usage and coverage. This section concerns the management of thesauri and attempts (1) to show that one problem with the previous efforts to maintain thesauri has been that they do not recognize and distinguish between the professional and the corporate perspective and (2) to sketch how thesaurus management can be done if this distinction is made. The previous efforts can be divided into three broad categories:

*The centrally driven approach*. The conventional approach to thesaurus maintenance, in practice as well as in the literature, consists in manual development of a number of successive versions, see for example Soergel (1974) and Batty (1989). Suggestions for modifications are collected over a period of time, then the responsibility of bringing the thesaurus up to date is given to a specialized unit—a thesaurus group. A centralized thesaurus group is considered necessary to ensure quality and consistency, for instance by keeping the overall structure of the thesaurus in mind and remembering that removal of outdated material is as important as adding new. As an example the keyword list in the documentation system of Novo Nordisk is maintained by a specialized keyword group. Individual chemists can forward suggestions to this group; they can not make modifications themselves (Hertzum, 1993). The chemists often work with details and special cases and, thus, use a large number of very specific terms; in contrast, modifications of the keyword list are only carried out if they are considered to be of at least some general relevance.

*The automatic approaches*. In the literature much attention is devoted to attempts at turning thesaurus management into an automatic process. These efforts are all directed towards the development of one authoritative thesaurus version and follow three different passes: (1) to infer the modifications from the queries, see for example Güntzer et al. (1989); (2) to create thesauri from the text of the documents, see for example Crouch

(1990) and Crouch & Yang (1992); and (3) to merge existing thesauri, see for example Mili & Rada (1988) and McMath et al. (1989). However, the algorithms are highly speculative; for instance the merging algorithms rely on the assumption that conceptual relatedness between terms can be computed on the basis of their topological distance in the thesaurus. These approaches seem to need a quality-ensuring thesaurus group much more than modifications suggested by the end-users. So far it is an open question whether tools based on the automatic approaches can be of value to the thesaurus groups, specific needs of individual professionals are not considered at all.

*The end-user involving approaches*. A very simple way of involving the end-users directly in the maintenance process was early discussed by Reisner (1963; 1966). She essentially suggested to develop a growing thesaurus by asking the individual users to add terms and links whenever they considered it relevant. With this unmoderated approach the thesaurus was simply defined as the sum of the individual contributions. Later work has attached more importance to consistency and less to giving the users full control over the system. For example, Güntzer et al. (1989), among other things, enable the users to enter suggestions for insertions during their retrieval sessions. A term inserted by one user as, say, related to another term is marked as "suggested for insertion". The idea is that when other users subsequently access the same part of the thesaurus, the suggestion is displayed and they are asked to judge it. The judgements are logged and available to the thesaurus group. However, the users are expected to suggest modifications without having them available immediately afterwards, and if they make many suggestions they will frequently be asked to suspend their doings to make judgements.

The approach suggested here considers new, consistent versions and modifications wanted by individual end-users to be equally important. From the professional perspective there is a need for end-user modifiability; from the corporate perspective there is a need for versioning. Thus, two separate thesaurus facilities should be provided and they should be interfaced, for instance to allow local modifications to be up-loaded. Often, individual professionals use important terms long before they find their way into corporate thesauri. Enabling the professionals to enter such terms into their personal thesauri would also enable the thesaurus group to up-load these modifications and consider them the next time the corporate thesaurus is modified. With this approach the professionals are provided with a facility directed towards their needs rather than asked to make a sacrifice, as is almost the case in Güntzer et al. (1989). In the thesaurus group, the local modifications made by individual professionals and groups of professionals can be treated statistically, for example to extract the high-frequency modifications and discover parts of the thesaurus especially in need of updating.

To properly integrate end-user modifiability and versioning the thesaurus group should also be able to enforce certain constraints on the local modifications. These

constraints could, for instance, prevent everyone but the thesaurus group from modifying the top levels of the thesaurus and, thus, maintain it as one corporate thesaurus with a number of personal details. The constraints could also be used to grant some users the right to modify while others are allowed to view only, or to enforce more general rules such as ensuring that no user establishes one term as broader than another if the inverse relationship already exists. To support the development of thesauri with such facilities special purpose tools are needed, and according to Milstead (1990) current, commercial tools for thesaurus management are much more ordinary. The successful development of such a tool would enable different organizations to allow end-user modifiability to the extent they considered appropriate. Also, it would enable experimentation on how to support the thesaurus group and how to propagate new versions of the thesaurus to the users without discarding their personal modifications—problems I have been looking into but without getting through to usable solutions.

*Generic TSARS and general-purpose toolkits*

Individual users have different needs, organizations have different needs, and furthermore what appears to be needed at some point in time will inevitably be subject to subsequent modifications. TSARS should be designed with this in mind. Though the need for changeability is often underestimated or sacrificed in favour of e.g. response time requirements, a number of approaches to the development of modifiable systems have also evolved. One of these relies on the so-called generic systems and extends a standard systems approach with facilities for extensions and tailoring; another relies on general-purpose toolkits and adds generality and flexibility to a tailor-made systems approach.

*Generic TSARS.* Like standard systems, generic systems are application specific and include all the facilities needed to build a complete application. Unlike standard systems, generic systems include more facilities than needed in any single application, and from this pool a number of facilities are selected and set up to fit the different and changing needs of individual customers. To provide this pool of facilities generic systems embody all the work involved in modelling the application area in a way which both accords with practice and supports systems development. This sets generic systems principally apart from general-purpose toolkits.

As argued in Hertzum (1994a) generic systems have several advantages including that large portions of analysis, design, coding, and debugging are done only once; that potential customers can visit organizations in which the system is already operational; and that customers can use prototyping to reach their system. The disadvantages include a risk that the generic system frames too strongly the customer’s thinking about what is needed and a response time overhead due to the more general-purpose solutions utilized to achieve flexibility. The generic systems approach seems appropriate to numerous

applications, provided the market for them is sufficiently large to call for alternatives to individual, tailor-made systems and too varied to call for a standard system. Several TSARS applications fulfill these requirements; indeed a number of widely used TSARS are generic, e.g. Scanjour and BRS/Search.

*General-purpose toolkits*. The first computers were programmed directly in machine code, later a series of increasingly more powerful, general, and flexible tools have been developed, e.g. assemblers, general-purpose programming languages, and various general-purpose toolkits such as database management systems (DBMS). General-purpose toolkits are directed towards a problem common to many applications and attempt to solve it in a way sufficiently general and flexible to ensure that the toolkit is applicable to a large number of these applications. A toolkit makes up only part of an application, and therefore the ability to interact with programming languages, other toolkits and the like is essential. Often, toolkits comply to certain standards and thus ensure that the applications do so too, but otherwise the toolkits strive to leave the design of applications entirely to the systems developers. Thus, while generic systems focus on one application area and attempt to address all aspects of it, general-purpose toolkits focus on one aspect of an application and attempt to address all application areas involving it.

A prominent example of general-purpose toolkits is relational DBMS which have established themselves as a standard across a broad range of applications, but only with regard to the file handling aspects. Other tools are needed for the rest of the application, for example the user interface and the code which provides the application specific functionality. As argued in Hertzum (1994b) the changeability of relational DBMS seems to make them well-suited as the basis of a TSARS development environment unifying the text model, the hypertext model, and the relational model. Successful TSARS are often subject to manifold changes during their lifetime, and their continued success is largely due to the changes being incorporated into the original structure and idea of the system in a smooth way, see Naur (1985). Therefore Hertzum (1994b) considers the changeability of the relational model more intrinsic to a TSARS development environment than the specific text-related facilities of the text model and the hypertext model.

To some extent generic TSARS and general-purpose toolkits are competing products. Currently, an increasing number of organizations prefer generic systems developed by software houses to in-house development based on general-purpose toolkits (Bansler & Havn, 1994). A major consequence of this increase in the use of generic systems is that the systems development process is split in two, on the one hand the development and evolution of the generic system and on the other hand the organizational implementation and maintenance of the individual installations. In the development of the generic systems it seems natural to utilize general-purpose toolkits; afterwards what the customers buy is generic systems. Thus, generic TSARS and general-purpose toolkits are

both important to the development of future, flexible TSARS, but at different levels. This two-level approach is embodied in the unified TSARS development environment discussed in Hertzum (1994b) and also found in commercial systems such as Scanjour. Furthermore, the emphasis on modifiability which is inherent in this approach seems to make it suited for the development of TSARS adding a third level of flexibility, end-user modifiability.

# 6. Conclusion

Documentation work, here studied as an aspect of professional work, is an activity on which countless organizations spend a substantial amount of resources every year. Documentation work consists of filing, managing, and retrieving documents and other material pertaining to the primary work; in addition the preparation of the documentation is sometimes considered part of the documentation work. Despite repeated attempts to reduce documentation work to a mechanical or clerical task it is, to a certain extent, inseparable from the primary work. One reason for this is that, in some cases, the documentation work is intrinsic to the intellectual process through which the primary work is done; another reason is that documentation work is not an end in itself but a means to support the primary work.

The perception and organization of the documentation work is deeply affected by the nature of the primary work. For example, researchers are allowed more freedom in their primary work as well as their documentation work than civil servants working in much regulated areas. However, the primary work is also affected by the organization of the documentation work; an illustrative example is provided by the Work Injury Office. This mutual dependence emphasizes that the empirical basis of studies like this must be kept in mind. This study rests on work involving several kinds of professionals—chemists, civil servants, lawyers, and a group of semi-professionals, namely computer science students. These groups of professionals have first been studied separately, i.e. with attention to the different kinds of primary work. Afterwards, in this report, it has been attempted to establish a more general picture, but with much emphasis on diversity and frequent use of examples from the underlying studies.

### *6.1 Documentation work and its organization*

Large individual differences exist in the perception and organization of the documentation work, differences among professionals as well as among organizations. Nevertheless, the professionals share a basic approach which has been called the professional perspective and stands in sharp contrast to the corporate perspective common to management and organization-wide bodies such as journalization units.

The professional perspective on the documentation work is centered around the professionals' offices, woven into their primary work, and characterized by exploiting the

possibilities arising from knowing the documents involved. Typically, the documents are organized according to their relation to the professionals' primary work, for example documents pertaining to a certain project, and retrieval essentially consists in recalling this relation. The professional perspective is directed towards organizing the professionals' material for their own convenience, not to provide others with access to it.

The corporate perspective is intended to preserve internal documents and make them available to other people than the authors. Thus, the persons retrieving from the corporate archive can not be expected to possess knowledge about what documents the archive contains. This unknown-universe situation is characteristic of the corporate perspective and the reason why it primarily focuses on filing. Careful, systematic filing is a prerequisite for retrieval, because the information recorded during filing is the searcher's only source of information about the documents. In a known universe, e.g. in the professionals' offices, emphasis is usually on retrieval because it is considered preferable to transfer part of the time spent on documentation work from filing to retrieval which is immediately valuable.

Despite the fundamental differences between the professional and the corporate perspective they coexist. However, they largely do so by refraining from exploiting the indisputable overlap between the perspectives. Instead the professionals interact with their colleagues who, among other things, provide them with much of the information they would otherwise have to retrieve from the corporate archive. With respect to the organizations, it generally seems that the potential of achieving mutual advantages by interconnecting the perspectives more closely is not recognized or not considered worth pursuing.

The unconnectedness of the two perspectives is, for instance, illustrated by their minimal overlap in terms of the people responsible for doing the documentation activities. On the one hand, it is left entirely to the professionals to organize their own documents; on the other hand, the activities pertaining to the corporate perspective are usually delegated to secretaries and specialized units to the widest extent possible. However, as the professionals have become acquainted with computers through their adoption of text processing and as TSARS have been acquired, certain expectations have been attached to end-user retrieval. So far, these expectations have not been fulfilled; the professionals prefer to delegate retrieval. Four reasons for this have been identified:

- Systems overload. TSARS constitute an increase in the amount of information to be processed and may, thus, cause a variant of information overload, here termed systems overload. It is inconceivable to get acquainted with all the systems available and hard just to find out which would be worth approaching.
- Critical mass. When TSARS are introduced they contain no documents, and during a shorter or longer transitional period retrieval involves both searching in the old

system and in the new. Before the new system is considered the major one it must reach a certain size, its critical mass. Until that happens many will hesitate to use it.

- Manual retrieval often outperforms online retrieval. Online retrieval seems to be a demanding task which many users experience difficulties in dealing with. Thus, at present, i.e. with contemporary TSARS and professionals without much experience in using them, reluctance towards end-user retrieval might be cost-effective.
- A lengthy process. Efforts to provide computer support for documentation work often focus narrowly on automating previously manual routines in which end-user retrieval was not an option. This illustrates how difficult it is to transcend the existing situation and rethink it given new possibilities provided by, for instance, technology.

While end-user retrieval is about who performs the retrieval activities, end-user modifiability, i.e. enabling individual professionals or groups of cooperating professionals to modify the TSARS, is primarily about the possibility to make local modifications. Therefore, the present, limited adoption of end-user retrieval is peripheral to end-user modifiability—it is of little significance whether facilities for end-user modifiability are operated by professionals or secretaries. In keeping their documentation work in accordance with their primary work professionals often need to accompany changes in their primary work with modifications in the organization of their documentation work. This need for frequent, local modifications is not met by the corporate efforts to keep the TSARS up to date; usually these efforts consist in collecting suggestions for modifications over a period of time and then prepare a new version. As an example, many professionals know and use a number of terms long before they, perhaps, find their way into corporate thesauri. Consequently, either TSARS include facilities for end-user modifiability or the professionals are forced to discard the TSARS or adapt to them. It seems as if TSARS should supplement versioning with facilities for end-user modifiability to enable the incorporation of changes important to the professionals' primary work as soon as the professionals become aware of such changes.

### *6.2 Computer support for documentation work*

Many aspects of documentation work and its computer support are poorly understood and surrounded by much uncertainty and hesitation, in practice as well as in the literature. TSARS form a varied group of systems but nevertheless seem to share a number of facilities and requirements. Systems with a broad range of facilities and capable of being tailored to meet the diverse and changing needs of different customers appear a promising and cost effective way to develop many TSARS. If application-specific, such systems are called generic TSARS; with less emphasis on specific applications they could be called TSARS development environments. The notion of a TSARS development environment provides an inspiring background for the subsequent suggestions regarding computer

support for documentation work. These suggestions summarize the seemingly central problems and promising solutions identified in this study, and they point to issues which call for future work.

*TSARS based on relational databases*. The changeability of relational DBMS seems to make them well-suited not only as the basis of individual TSARS, but also as the basis of a TSARS development environment. As TSARS move into new application areas, the data models on which they are based are applied in other contexts than those in which the data models were originally devised. As a consequence, many applications do not fit a single data model, rather they need facilities from a combination of data models. Thus, to form a TSARS development environment the relational model should be extended with certain facilities known from data models developed specifically for TSARS, e.g. the text model and the hypertext model. In addition, this study hypothesizes that the facilities discussed below are important to a broad range of TSARS and, thus, should be supported by a TSARS development environment.

Karnov's Law Database and TeSS are both examples of TSARS based on relational DBMS, and both seem to support the idea that the relational model has potential as the basis of TSARS. However, future work is needed to develop a TSARS development environment based on the relational model and test it in practice. Especially, long, empirical studies are needed to investigate the changes to which TSARS are subjected and the adequacy of the facilities intended to support modifications.

*A Janus-faced approach to TSARS development*. The interplay between the professional and the corporate perspective is in this study considered of strategic importance and, thus, central to success in providing computer support for documentation work. Emphasizing this interplay brings into focus the need for facilities, work routines and the like which contribute to bridge the gap between the perspectives. To bridge this gap and ensure that both perspectives are acknowledged, it is suggested to design TSARS as two separate but interfaced systems. One system should support the professional perspective, the other the corporate perspective. The interface should enable information made available by the professionals to be up-loaded to the corporate part of the system and enable individual professionals to down-load information from the corporate part.

While the existence of the professional and the corporate perspective is derived from empirical studies, the importance of facilities bridging the gap between them is a hypothesis. There is a large need for future work which examines this hypothesis, i.e. for the development of prototype TSARS based on the Janus-faced approach and empirical evaluations of their use in challenging, organizational settings. Some of these studies should go beyond the archive orientation which characterizes most TSARS and attempt to view documentation work in the context of e.g. project management. TSARS already

handle documents and record document flow and, therefore, seem suited for development into or integration with so-called work flow tools.

*Background filing*. Organizations are dependent on the filing activities, and the more TSARS get involved in new tasks such as project management, the larger the dependence will get. From the corporate perspective the only way to ensure that the relevant documents are filed is to demand that everything is filed. From the professional perspective this is a time-consuming and for many documents pointless task. To generate and maintain the care necessary in filing the more important documents it is essential not to flood the professionals with regulations concerning the less important ones. Background filing consists in an automatic base registration of all online documents together with the assumption that this base registration is sufficiently good to render human involvement in the filing of the less important documents superfluous. In addition to the document contents a background filing facility can provide certain attributal information, such as author, date, and automatically determined keywords. However, an essential aspect of background filing is the establishment of an accompanying routine for selecting the important documents and subjecting them to human indexing.

Currently, the potential of background filing is just a hypothesis. Technically much can be achieved simply by combining a backup system and an information retrieval system. In many cases, the type of the documents can be used to determine whether they are important. This information could, for example, be made available to the system through the use of standard document setups. Furthermore, markup languages such as SGML can be used to extend the documents with information about their structure and, thereby, enable automatic extraction of title, reference number, table of contents and the like. In addition to the technical aspects there is a definite need for future work addressing the organizational implementation of background filing. Such work should establish whether or not actual advantages can be gained, assess the concrete organizational impact of background filing facilities, develop work routines around them and so on.

*Combining local modifications and versioning*. While background filing is primarily directed towards the corporate perspective, facilities for combining local modifications and versioning are grounded in the recognition of the coexistence and importance of the professional and the corporate perspective. If such facilities are not available, then either any local modifications are lost when a new version is installed or the installation of new versions is avoided to preserve local modifications. It should be noted that fully automatic solutions to the problem do not exist because a significant part of it consists, not in deciding how to combine two modifications, but in deciding which one of two conflicting modifications the individual user prefers.

As of yet the problem is poorly understood. Future work is needed to devise facilities supporting the combination of local modifications and versioning and to find out

how much attention different groups of professionals are prepared to spend on keeping their documentation tools up to date. The problem appears in a number of contexts, for instance thesaurus management and version management of application systems, and this study hypothesizes that solutions are required to make end-user modifiability an attractive supplement to versions developed centrally.

*Integration of browsing and querying*. Technically, it is straightforward to provide TSARS with facilities for both browsing and querying, but simply having both available does not seem to be of advantage to the users. Rather, they seem to be exposed to an overload very similar to the systems overload which, along with other things, causes professionals to avoid end-user retrieval. However, browsing and querying seem, to a certain extent, suited to different types of searches, and thus some TSARS need to support both.

Future work is needed to generate and test ideas which allow these two retrieval techniques to be combined without degrading the users' performance. Presumably, the best way to combine browsing and querying depends on the nature of the primary work, on the individual users, as well as on various other factors. One issue which needs attention is the nature, size, and handling of individual differences among the users. The average performance often hides substantial differences among individual users, and little is known about how these differences are connected to properties of TSARS, personal characteristics and so on. Some users might perform best if the profiles of the browsing and querying facilities are sharp and intended to accentuate their capabilities and differences; others if the facilities are built tightly together and the issue of when to browse and when to query is more or less dissolved.

During the past 30 years computers have been assigned a key role in the efforts to provide support for documentation work. Progress has been made, but we also, repeatedly, get new evidence, such as the TeSS experiment, that computer-held information does not outperform 'old-fashioned' hard copies. It has proved to be inherently difficult to turn the potential of computers into TSARS which enable the users to file, manage, and retrieve their documentation in an efficient and useful way.

# References


Attewell, P. & Rule, J. (1984). Computing and organizations: what we know and what we don't know. *Communications of the ACM*, **27**(12), 1184-1192.

Bansler, J. P. & Havn, E. (1994). Information systems development with generic systems. In W. R. J. Baets, ed., *Proceedings of the Second European Conference on Information Systems* (Breukelen, The Netherlands, May 30-31, 1994). pp. 707-715.

Batty, D. (1989). Thesaurus construction and maintenance: a survival kit. *Database*, **12**(1), 13-20.

Blair, D. C. & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document-retrieval system. *Communications of the ACM*, **28**(3), 289-299.

Borgman, C. L. (1989). All users of information retrieval systems are not created equal: an exploration into individual differences. *Information Processing & Management*, **25**(3), 237-251.

Brancheau, J. C. & Brown, C. V. (1993). The management of end-user computing: status and directions. *ACM Computing Surveys*, **25**(4), 437-482.

Broløs, J. H., Frøkjær, E., Hertzum, M., Lárusdóttir, M. K., Pilgaard, K. B. & Sørensen, F. S. (1993). *TeSS-projektet. Udvikling af et system til eksperimentel undersøgelse af brugergrænseflader til edb-baseret tekstsøgning*. DIKU report 93/18. DIKU, Copenhagen.

Bush, V. (1945). As we may think. *Atlantic Monthly*, **176**(1), 101-108.

Crouch, C. (1990). An approach to the automatic construction of global thesauri. *Information Processing & Management*, **26**(5), 629-640.

Crouch, C. J. & Yang, B. (1992). Experiments in automatic statistical thesaurus construction. In N. Belkin, P. Ingwersen & A. M. Pejtersen, eds., *Proceedings of the Fiftenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Copenhagen, Denmark, June 21-24, 1992). pp. 77-88. ACM Press, New York.

Diaper, D., ed. (1989). *Task analysis for human-computer interaction*. Ellis Horwood, Chichester.

Dreyfus, H. L. & Dreyfus, S. E. (1986). *Mind over machine. The power of human intuition and expertise in the era of the computer*. Basil Blackwell, Oxford.

Egan, D. E. (1988). Individual differences in human-computer interaction. In M. Helander, ed., *Handbook of Human-Computer Interaction*. pp. 543-568. Elsevier Science Publishers, North-Holland.

Ellis, D. (1989). A behavioural approach to information retrieval system design. J*ournal of Documentation*, **45**(3), 171-212.

Ellis, D., Cox, D. & Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation*, **49**(4), 356-369.

Engelbart, D. C. & English, W. K. (1968). A research center for augmenting human intellect. *AFIPS Conference Proceedings*, vol. 33, part 1. pp. 395-410. Thompson Book Company, Washington, D.C.

Fischer, G. & Girgensohn, A. (1990). End-user modifiability in design environments. In J. C. Chew & J. Whiteside, eds., *Empowering People. CHI'90 Conference Proceedings* (Seattle, Washington, April 1-5, 1990). pp. 183-191. ACM Press, New York.

Foskett, D. J. (1985). Thesaurus. In E. D. Dym, ed., *Subject and Information Analysis*. pp. 270-316. Marcel Dekker, New York.

Güntzer, U., Jüttner, G., Seegmüller, G. & Sarre, F. (1989). Automatic thesaurus construction by machine learning from retrieval sessions. *Information Processing & Management*, **25**(3), 265-273.

Halasz, F. G. (1988). Reflections on Notecards: seven issues for the next generation of hypermedia systems. *Communications of the ACM*, **31**(7), 836-852.

Hertzum, M. (1993). Information retrieval in a work setting: a case study of the documentation part of chemists' work. In J. P. Bansler, K. Bødker, F. Kensing, J. Nørbjerg & J. Pries-Heje, eds., *Proceedings of the 16th IRIS. Information Systems Research Seminar in Scandinavia* (Copenhagen, August 7-10, 1993). pp. 786-798. DIKU report 93/16. DIKU, Copenhagen. Also available as appendix A in this report.

Hertzum, M. (1994a). Computer support for journalization in the Danish central government. Submitted for publication. Available as appendix B in this report.

Hertzum, M. (1994b). A comparison of three data models for text storage and retrieval systems: the relational model revisited. In M. Khosrowpour, ed., *Managing Social and Economic Change with Information Technology. Proceedings of the 5th International Conference of the Information Resources Management Association* (San Antonio, Texas, May 22-25, 1994). pp. 448-455. Idea Group Publishing, Harrisburg. Also available as appendix C in this report.

Hertzum, M. & Frøkjær, E. (1994). Browsing and querying in online documentation: a study of user interfaces and the interaction process. Submitted for publication. Available as appendix D in this report.

Hertzum, M., Søes, H. & Frøkjær, E. (1993). Information retrieval systems for professionals: a case study of computer supported legal research. *European Journal of Information Systems*, **2**(4), 296-303. Also available as appendix E in this report.

Jørgensen, A. H. & Sauer, A. (1990). The personal touch: a study of users' customization practice. In D. Diaper, D. Gilmore, G. Cockton & B. Shackel, eds., *Proceedings of INTERACT'90. Third IFIP Conference on Human-Computer Interaction*. pp. 561-565. North-Holland, Amsterdam.

Kay, A. & Goldberg, A. (1977). Personal dynamic media. *IEEE Computer*, **10**(3), 31-41.

Lancaster, F. W., Elliker, C. & Connell, T. H. (1989). Subject analysis. *Annual Review of Information Science and Technology*, **24**, 35-84.

Leavitt, H. J. (1964). Applied organizational change in industry: structural, technical, and human approaches. In W. W. Cooper & H. J. Leavitt, eds., *New Perspectives in Organization Research*. pp. 55-71. New York.

Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics*, **1**(1), 4-11.

Mander, R., Salomon, G. & Wong, Y. Y. (1992). A 'pile' metaphor for supporting casual organization of information. In P. Bauersfeld, J. Bennett & G. Lynch, eds., *CHI'92 Conference Proceedings* (Monterey, California, May 3-7, 1992). pp. 627-634. ACM Press, New York.

Marchionini, G. (1989). Making the transition from print to electronic encyclopaedias: adaptation of mental models. *International Journal of Man-Machine Studies*, **30**(6), 591-618.

Mason, M. V. (1986). Adaptive command prompting in an on-line documentation system. *International Journal of Man-Machine Studies*, **25**(1), 33-51.

McLean, E. R., Kappelman, L. A. & Thompson, J. P. (1993). Converging end-user and corporate computing. *Communications of the ACM*, **36**(12), 79-92.

McMath, C. F., Tamaru, R. S. & Rada, R. (1989). A graphical thesaurus-based information retrieval system. *International Journal of Man-Machine Studies*, **31**(2), 121-147.

Meadow, C. T. (1992). *Text information retrieval systems*. Academic Press, San Diego, California.

Mili, H. & Rada, R. (1988). Merging thesauri: principles and evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10**(2), 204-220.

Milstead, J. L. (1990). Thesaurus software packages. In D. Henderson, ed., *ASIS'90. Proceedings of the 53rd ASIS Annual Meeting* (Toronto, Ontario, November 4-8, 1990). pp. 3-15. Learned Information, Medford, New Jersey.

Mintzberg, H. (1983). *Structures in fives: designing effective organizations*. Prentice-Hall, Englewood Cliffs, New Jersey.

Mischo, W. H. & Lee, J. (1987). End-user searching of bibliographic databases. *Annual Review of Information Science and Technology*, **22**, 227-263.

Naur, P. (1965). The place of programming in a world of problems, tools, and people. *Proceedings of the IFIP Congress 65*. pp. 195-199. Also published as section 1.1 in P. Naur, *Computing: A Human Activity*, ACM Press/Addison-Wesley, New York, 1992.

Naur, P. (1985). Programming as theory building. *Microprocessing and Microprogramming*, **15**(5), 253-261. Also published as section 1.4 in P. Naur, *Computing: A Human Activity*, ACM Press/Addison-Wesley, New York, 1992.

Naur, P. (1988). Human knowing, language, and discrete structures. Published as section 8.7 in P. Naur, *Computing: A Human Activity*, ACM Press/Addison-Wesley, New York, 1992.

Naur, P. (1990). Computing and the so-called foundations of the so-called sciences. Presented at *Informatics Curricula for the 1990s*, IFIP Working Group 3.2 Workshop (Providence, Rhode Island, April 6, 1990). Also published as section 1.5 in P. Naur, *Computing: A Human Activity*, ACM Press/Addison-Wesley, New York, 1992.

Reisner, P. (1963). Construction of a growing thesaurus by conversational interaction in a man-machine system. *Proceedings of the American Documentation Institute, 26th Annual Meeting* (Chicago, Illinois, October, 1963). pp. 99-100.

Reisner, P. (1966). Evaluation of a 'growing' thesaurus. Presented at *the Center for the Information Sciences, Seminar in the Information Sciences* (Lehigh University, Bethlehem, Pennsylvania, March 10, 1966). Available as Research Paper RC-1662, August 9, 1966, IBM Watson Research Center, Yorktown Heights, New York.

Salton, G., ed. (1971). *The Smart retrieval system—experiments in automatic document processing*. Prentice Hall, Englewood Cliffs, New Jersey.

Salton, G. (1986). Another look at automatic text-retrieval systems. *Communications of the ACM*, **29**(7), 648-656.

Salton, G. & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, **24**(5), 513-523.

Soergel, D. (1974). *Indexing languages and thesauri: construction and maintenance*. Wiley, Los Angeles, California.

Sommerville, I. (1992). *Software engineering*. 4th edition. Addison-Wesley, Reading, Massachusetts.

Tague, J. M. (1981). User-responsive subject control in bibliographic retrieval systems. *Information Processing & Management*, **17**(3), 149-159.

Tenopir, C. (1984). Full-text databases. *Annual Review of Information Science and Technology*, **19**, 215-246.

Tenopir, C. & Ro, J. S. (1990). *Full text databases*. Greenwood Press, New York.

Tyler, S. W. & Treu, S. (1989). An interface architecture to provide adaptive task-specific context for the user. *International Journal of Man-Machine Studies*, **30**(3), 303-327.

# Dansk resumé

Denne rapport er indleveret til bedømmelse som PhD-afhandling på DIKU, Datalogisk Institut ved Københavns Universitet. Den afslutter arbejde udført fra januar 1992 til august 1994. Mit PhD-studium er dokumenteret i fem artikler, en teknisk rapport og denne rapport, som er sammenfatningen.

Studiet handler om dokumentationsarbejde, og hvordan det bedst støttes med edb i form af såkaldte dokumentationssystemer. Dokumentationsarbejde, der her studeres som et aspekt af fagfolks arbejde, består af arkivering, håndtering og fremfinding af dokumenter og andet materiale med tilknytning til fagfolkenes primære arbejde. Endvidere betragtes udarbejdelsen af dokumentationen somme tider, men ikke her, som en del af dokumentationsarbejdet. Til trods for gentagne forsøg på at reducere dokumentationsarbejde til en mekanisk, rutinepræget opgave er det i et vist omfang uadskilleligt fra fagfolkenes primære arbejde. En årsag til dette er, at dokumentationsarbejdet i visse tilfælde er en central del af den intellektuelle proces, som driver det primære arbejde; en anden årsag er, at dokumentationsarbejde ikke er et mål i sig selv, men et middel til at støtte det primære arbejde.

Arten af fagfolkenes primære arbejde har afgørende indflydelse på opfattelsen og organiseringen af dokumentationsarbejdet; men det primære arbejde påvirkes også af, hvordan dokumentationsarbejdet er organiseret. Denne gensidige afhængighed understreger vigtigheden af at holde sig grundlaget for studier som dette for øje. Dette studium er baseret på empirisk og eksperimentelt arbejde, der involverer flere slags fagfolk - kemikere, sagsbehandlere, jurister og en gruppe semi-fagfolk, i form af datalogistuderende. Formålet med studiet er:

- at studere den praktiske, organisatoriske kontekst, dokumentationsarbejde udspiller sig i, med henblik på at afdække dets art og organisering,
- at fremhæve forandring og forskelligartethed som essentielle egenskaber ved fagfolks arbejde og foreslå visse konsekvenser for dokumentationssystemer,
- at evaluere forskellige brugergrænseflader til et sådant system med hensyn til deres betydning for kvaliteten og effektiviteten af dokumentationsarbejdet.

*Dokumentationsarbejde og dets organisering*

Dokumentationsarbejde er en aktivitet som utallige organisationer hvert år bruger betragtelige ressourcer på. Der er store individuelle forskelle i opfattelsen og organise-

ringen af dokumentationsarbejdet, mellem forskellige fagfolk såvel som mellem forskellige organisationer. Alligevel er fagfolkene fælles om en grundlæggende tilgang, det professionelle perspektiv, der står i skarp kontrast til det virksomhedsperspektiv, som er fælles for ledelsen, journalenheder og lignende.

Det professionelle perspektiv på dokumentationsarbejdet er centreret om fagfolkenes kontorer, vævet ind i deres primære arbejde og karakteriseret ved at udnytte de muligheder, der ligger i at kende de involverede dokumenter. Typisk er dokumenterne organiseret på basis af deres relation til det primære arbejde, fx dokumenter knyttet til et bestemt projekt, og fremfinding består først og fremmest i at genkalde denne relation. Det professionelle perspektiv er rettet mod at give fagfolkene bekvem adgang til deres egne dokumenter, ikke mod at give andre adgang til dem.

Virksomhedsperspektivet har til formål at sikre, at interne dokumenter bevares og gøres tilgængelige for andre end forfatterne. Personer, der søger i en organisations arkiv, kan derfor ikke forventes at vide hvilke dokumenter, arkivet indeholder. Det er karakteristisk for virksomhedsperspektivet, at søgning foregår i et ukendt univers, og af den grund ofres der stor opmærksomhed på arkivering. Omhyggelig, systematisk arkivering er en forudsætning for fremfinding, fordi den information, der registreres under arkiveringen, er de søgendes eneste kilde til information om dokumenterne. I et kendt univers, fx på fagfolkenes kontorer, er hovedvægten sædvanligvis på fremfinding, da det foretrækkes at flytte noget af den tid, der bruges på dokumentationsarbejdet, fra arkivering til fremfinding, som er en umiddelbart nyttig aktivitet.

Til trods for de fundamentale forskelle mellem de to perspektiver eksisterer de side om side. De udnytter imidlertid så godt som ikke det indiskutable overlap mellem perspektiverne. I stedet holder fagfolkene sig i kontakt med deres kolleger, som blandt andet giver dem meget af den information, de ellers måtte søge i organisationens arkiv. Med hensyn til organisationerne ser det generelt ud til, at fordelene ved at sammenkæde perspektiverne ikke er erkendt eller ikke anses for værd at forfølge.

Fraværet af en sådan sammenkædning illustreres blandt andet af det minimale overlap med hensyn til hvem, der forestår dokumentationsaktiviteterne. På den ene side er det overladt til fagfolkene at organisere deres egne dokumenter; på den anden side er de aktiviteter, der er knyttet til virksomhedsperspektivet, som regel i videst muligt omfang delegeret til sekretærer og specialiserede enheder. Efterhånden som fagfolkene er blevet fortrolige med edb gennem deres tilegnelse af tekstbehandling, og efterhånden som dokumentationssystemer er blevet erhvervet, er der imidlertid blevet knyttet visse forventninger til slutbruger-fremfinding. Indtil videre er disse forventninger ikke blevet opfyldt; fagfolkene foretrækker delegeret fremfinding. Fire årsager til dette er blevet identificeret:

- Systemoversvømmelse. Dokumentationssystemer udgør i sig selv en forøgelse af den mængde information, fagfolkene skal behandle, og kan forårsage en variant af

informationsoversvømmelse, her kaldet systemoversvømmelse. Det er udelukket at blive fortrolig med alle de tilgængelige systemer og svært blot at finde ud af, hvilke det kunne betale sig at se nærmere på.

- Kritisk masse. Når dokumentationssystemer introduceres, indeholder de ingen dokumenter, og gennem en kortere eller længere overgangsperiode omfatter fremfinding både søgning i det gamle og det nye system. Før det nye system betragtes som det primære, må det nå en vis størrelse, dets kritiske masse. Indtil det sker, vil mange tøve med at bruge det.
- Manuel fremfinding er ofte væsentligt mere effektivt end edb-baseret. Edb-baseret fremfinding ser ud til at være en krævende opgave, som mange brugere har svært ved at håndtere. I øjeblikket, dvs med de eksisterende dokumentationssystemer og med fagfolk uden megen erfaring i at bruge dem, kan modvillighed overfor slutbruger-fremfinding således være rationelt.
- En langstrakt proces. Bestræbelser på at tilvejebringe edb-støtte til dokumentationsarbejde fokuserer ofte snævert på at automatisere tidligere manuelle rutiner, hvor slutbruger-fremfinding ikke var muligt. Det illustrerer, hvor svært det er at overskride den eksisterende situation og gentænke den på basis af nye muligheder skabt af fx ny teknologi.

Mens slutbruger-fremfinding drejer sig om, hvem der udfører fremfindingsaktiviteterne, drejer slutbruger-foranderlighed sig primært om muligheden for at lave lokale modifikationer. Den øjeblikkelige, begrænsede udbredelse af slutbruger-fremfinding er derfor kun af perifer betydning i forhold til slutbruger-foranderlighed - det er ikke afgørende om faciliteter til slutbruger-foranderlighed betjenes af fagfolkene selv eller sekretærer. I bestræbelserne på at holde deres dokumentationsarbejde i kontakt med det primære arbejde har fagfolk ofte brug for at ledsage ændringer i det primære arbejde af modifikationer i organiseringen af dokumentationsarbejdet. Dette behov for hyppige, lokale modifikationer opfyldes ikke af de organisatoriske bestræbelser på at holde dokumentationssystemerne à jour; disse bestræbelser består sædvanligvis i at indsamle forslag til ændringer over et stykke tid og derefter lave en ny version. Eksempelvis kender og bruger fagfolk mange termer, længe før de, muligvis, optages i organisationens tesaurus. Konsekvensen af dette er, at hvis dokumentationssystemer ikke indeholder faciliteter til slutbruger-foranderlighed, er fagfolkene tvunget til at se bort fra systemerne eller tilpasse sig til dem. Dokumentationssystemer bør derfor supplere faciliteter til versionsudvikling med faciliteter til slutbruger-foranderlighed for at muliggøre, at ændringer af betydning for fagfolkenes primære arbejde kan indarbejdes, så snart fagfolkene bliver klar over dem.

*Edb-støtte til dokumentationsarbejde*

Mange aspekter af dokumentationsarbejde og dets edb-støtte er mangelfuldt forstået og omgivet af usikkerhed og tøven, i praksis såvel som i litteraturen. Dokumentationssystemer udgør en varieret gruppe af systemer, men har alligevel en række fælles egenskaber. Systemer med en bred vifte af faciliteter og mulighed for at blive skræddersyet til individuelle kunders forskelligartede og foranderlige behov forekommer at være en lovende måde at udvikle mange dokumentationssystemer på. Hvis sådanne systemer er applikationsspecifikke, kaldes de generiske dokumentationssystemer; uden dette fokus på specifikke applikationer kunne de kaldes udviklingsmiljøer for dokumentationssystemer. Ideen om et udviklingsmiljø for dokumentationssystemer udgør en inspirerende baggrund for de følgende forslag angående edb-støtte til dokumentationsarbejde. Disse forslag sammenfatter de i dette studium fundne centrale problemer og lovende løsninger, og de peger på områder og emner for videre arbejde.

*Dokumentationssystemer baseret på relationelle databaser*. Relationelle databasestyresystemers fleksibilitet forekommer at gøre dem velegnede, ikke blot som basis for enkeltstående dokumentationssystemer, men også som grundlag for et udviklingsmiljø for sådanne systemer. Efterhånden som dokumentationssystemer breder sig til nye applikationsområder, anvendes de datamodeller, systemerne er baseret på, i andre sammenhænge end dem, de oprindeligt blev udviklet til. Derfor passer mange applikationer ikke til en enkelt datamodel, men har behov for faciliteter knyttet til en kombination af datamodeller. For at basere et udviklingsmiljø på den relationelle model skulle den altså suppleres med visse faciliteter fra datamodeller, der er udviklet specielt med henblik på dokumentationssystemer, fx tekstmodellen og hypertekstmodellen. Derudover fremstår de faciliteter, der diskuteres nedenfor, i dette studie som relevante for mange dokumentationssystemer, og de bør derfor støttes af et udviklingsmiljø.

Studiet omfatter udvikling af to eksempler på dokumentationssystemer baseret på relationelle databaser, og begge ser ud til at støtte ideen om, at den relationelle model har potentiale i denne sammenhæng. Der er imidlertid behov for videre arbejde med hensyn til at lave et udviklingsmiljø baseret på den relationelle model og afprøve det i praksis. Specielt er der behov for længerevarende, empiriske studier af de ændringer, dokumentationssystemer gennemgår, og af styrker og svagheder ved de faciliteter, der sigter på at støtte modifikationer.

*En janushoved-tilgang til udvikling af dokumentationssystemer*. Samspillet mellem det professionelle perspektiv og virksomhedsperspektivet betragtes her som værende af strategisk betydning og dermed centralt for edb-støtte til dokumentationsarbejde. Ved at fremhæve dette samspil bringes behovet for faciliteter og arbejdsrutiner, der bidrager til at sammenkæde perspektiverne, i fokus. For at sammenkæde perspektiverne og sikre, at de begge anerkendes under systemudviklingen, foreslås det at designe dokumentations-

systemer som to separate, men forbundne, systemer. Det ene system skal støtte det professionelle perspektiv, det andet virksomhedsperspektivet. Forbindelsen mellem dem skal muliggøre, at information, som er gjort tilgængelig af fagfolkene, suges op i systemets organisationsdel, og at individuelle fagfolk kan trække information ned fra organisationsdelen.

Mens eksistensen af det professionelle perspektiv og virksomhedsperspektivet er udledt af empiriske studier, er vigtigheden af faciliteter, som sammenkæder perspektiverne, kun en hypotese. Der er et stort behov for videre arbejde, som tager denne hypotese op, dvs for udvikling af dokumentationssystemer baseret på janushoved-tilgangen og for empiriske evalueringer af deres brug i krævende, organisatoriske sammenhænge. Nogle af disse studier bør overskride den arkiv-orientering, som kendetegner de fleste dokumentationssystemer, og eksempelvis betragte dokumentations-arbejde i forbindelse med projektstyring. Dokumentationssystemer håndterer allerede dokumenter og registrerer dokumentbevægelser; de forekommer derfor egnede til også at indgå i håndteringen og styringen af fagfolks arbejdsopgaver

*Baggrundsarkivering*. Organisationer er afhængige af arkiveringsaktiviteterne, og jo mere dokumentationssystemerne involveres i nye opgaver som projektstyring, jo større vil denne afhængighed blive. Set fra virksomhedsperspektivet er det kun ved at kræve, at alt arkiveres, det kan sikres, at alt relevant materiale arkiveres. Set fra det professionelle perspektiv er det en tidskrævende og for mange dokumenter meningsløs opgave. For at generere og fastholde den nødvendige omhu i arkiveringen af de vigtige dokumenter er det afgørende, at fagfolkene ikke overbebyrdes med krav vedrørende de mindre vigtige dokumenter. Baggrundsarkivering består i en automatisk grundregistrering af alle elektroniske dokumenter kombineret med den antagelse, at denne grundregistrering er tilstrækkelig god til at overflødiggøre manuel behandling af de mindre vigtige dokumenter. Udover dokumenternes indhold kan en baggrundsarkiveringsfacilitet tilvejebringe visse oplysninger om dokumenterne, fx forfatter, dato og automatisk bestemte nøgleord. Et væsentligt aspekt ved baggrundsarkivering er imidlertid etableringen af en sideløbende rutine for udvælgelse af de vigtige dokumenter til manuel indeksering.

I øjeblikket er mulighederne i baggrundsarkivering kun en hypotese. Teknisk set kan meget opnås blot ved at kombinere et backup-system og et informationssøgesystem. Derudover kan opmærkningssprog såsom SGML bruges til at udvide dokumenterne med oplysninger om deres struktur og derved muliggøre automatisk udtræk af titel, journal-nummer, indholdsfortegnelse og lignende. Udover de tekniske aspekter er der et udtalt behov for videre arbejde med hensyn til den organisatoriske implementering. Sådant arbejde skulle fastslå, hvorvidt der kan opnås egentlige fordele ved baggrundsarkivering,

evaluere den konkrete organisatoriske betydning af sådanne faciliteter, udvikle arbejdsrutiner omkring dem og lignende.

*Kombination af lokale modifikationer og udarbejdelse af versioner*. Mens baggrundsarkivering primært er rettet mod virksomhedsperspektivet, udspringer faciliteter til at kombinere lokale modifikationer og versionsudvikling af en anerkendelse af den samtidige tilstedeværelse og vigtighed af det professionelle perspektiv og virksomheds-perspektivet. Hvis sådanne faciliteter ikke er til stede, går eventuelle lokale modifikationer tabt ved installation af en ny version, eller installation af nye versioner undgås for at bevare lokale modifikationer. Det skal bemærkes, at helt automatiske løsninger på problemet ikke eksisterer, da en væsentlig del af det består i at afgøre, hvilken af to uforenelige modifikationer brugeren foretrækker - ikke i at beslutte hvordan modifikationer skal kombineres.

På nuværende tidspunkt er problemet mangelfuldt forstået. Videre arbejde må anvise faciliteter, der støtter kombinationen af lokale modifikationer og versionsudvikling, og undersøge, hvor meget opmærksomhed forskellige grupper af fagfolk er villige til at ofre på at holde deres dokumentationssystem à jour. Problemet forekommer i en række sammenhænge, og i dette studie betragtes løsninger på det som en nødvendighed for at gøre slutbruger-foranderlighed til et attraktivt supplement til centralt udviklede versioner.

*Integration af browsing og søgning ved forespørgsler*. Teknisk set er det enkelt at udstyre dokumentationssystemer med faciliteter til både browsing og søgning ved forespørgsler; men blot det at have begge til rådighed ser ikke ud til at være en fordel for brugerne. De ser snarere ud til at blive udsat for noget meget lig den system-oversvømmelse, som er en af årsagerne til, at fagfolk undgår slutbruger-fremfinding. Browsing og søgning ved forespørgsler ser imidlertid ud til, i en vis udstrækning, at være egnet til forskellige typer af søgninger, og nogle dokumentationssystemer vil derfor skulle støtte dem begge.

Der er behov for videre arbejde med ideer, som tillader, at de to typer søge-teknikker kombineres, uden at brugerne derved hæmmes i deres arbejde. Den bedste måde at kombinere browsing og søgning ved forespørgsler på afhænger formodentlig af arten af det primære arbejde, de enkelte brugere og en række andre faktorer. Et område, hvor der er behov for videre arbejde, er med hensyn til arten, omfanget og håndteringen af individuelle forskelle mellem brugerne. Den gennemsnitlige brugseffektivitet skjuler ofte anseelige forskelle mellem de enkelte brugere, og vi ved kun lidt om, hvordan disse forskelle er knyttet til egenskaber ved dokumentationssystemerne, personlige karakteristika og lignende. Nogle brugere vil muligvis arbejde mest effektivt, hvis faciliteterne til browsing og søgning ved forespørgsler står skarpt overfor hinanden og derved tydeliggør styrker og forskelle; andre hvis faciliteterne bygges tæt sammen i et

forsøg på at guide brugeren og mere eller mindre opløse spørgsmålet om, hvornår der skal browses, og hvornår der skal forespørges.

Gennem de sidste 30 år har datamaskiner haft en central rolle i bestræbelserne på at støtte dokumentationsarbejde. Der er sket fremskridt; men vi ser også, igen og igen, beviser på, at elektronisk information ikke er 'gammeldags', trykt information overlegen. Det har vist sig grundlæggende svært at udfolde datamaskiners potentiale i dokumentationssystemer, som gør brugerne i stand til at arkivere, håndtere og fremfinde dokumentation på en effektiv og nyttig måde.

# Appendixes