Annelise Mark Pejtersen

Cognitive Engineering in Information Retrieval Domains – Merging Paradigms?

Introduction

For design of decision support systems for the staff involved in modern work environments it is necessary to consider carefully the changing role of human agents in work. Human agents are present in work systems for basically two different reasons: They are very flexible and versatile manipulators of objects and tools and, therefore, able to act as multi-purpose devices in complex work environments. In addition, they are adaptive problem solvers, able to generate new activity patterns on demand when required by changes in work requirements. All work domains involve both capabilities, but the technological development through the latest decades is rapidly changing the significance of the two capabilities. This change has great impacts on the proper design of support systems and also on the way human activity should be analysed and modelled for the design of intelligent information systems. Therefore, a new method for analysis and modelling of human activity called "Cognitive Engineering" has been emerging during the latest years which has been successfully merged with the Information Science disciplines during the design of the Book House system for information retrieval.

Merging Paradigms?

Two different approaches have been taken to the problem of collecting, storing, and retrieving information within two different research traditions based on different paradigms. One is found within information and library sciences, another within cognitive engineering. Due to the present trend toward 'integrated work stations,' however, time has now come to consider the benefit to be expected by merging the paradigms, research efforts, and the results already available (Pejtersen 1992a). In the present contribution, an attempt is made to illustrate the cognitive engineering approach to analysis of work domains and design of work stations and to demonstrate the need for a closer interaction with the information and library sciences.

Two Approaches to Information Retrieval

The information and library sciences and the cognitive engineering approaches to information indexing and retrieval, originating in the different work situations that have been the underlying precondition for the studies, have very distinct characteristics.

Library Sciences

A classic approach evolved from library sciences. The basic problem in this approach has been retrieval of documents from a large, more or less uniform collection in response to isolated questions posed by a user. The aim has been systematic, exclusive classification systems which made an unambiguous positioning and retrieval of the individual item of the collection possible in catalogues and on shelves. The sources of information in a library system originally were rather uniform sets of documents and standardised methods for identification of the individual items were established. For *fact literature* hierarchical, exclusive classification schemes evolved based on an analysis of the substance matter content of the documents. The indexing of documents became the craft of librarians who also served as mediators toward the queries of the users. For *fiction literature* the positioning and retrieval of documents normally were based on alphabetical indexing of authors. In both cases, the indexing and retrieval strategies were based on an analysis of the *documents,* with little consideration of the user or the problem that caused the user to visit the library, hence the need for mediation by the professional librarian.

The field of information science emerged as an applied science in the fifties along with the introduction of computer technology among scientists from many different scientific disciplines, whose expertise was organisation, management and retrieval of scientific information to other scientists within their own professional domain. The core of information science research is concerned with development of theories about the basic nature, definitions and boundaries of information. And it is concerned with the development of concepts and methods applicable to typical activities in relation to generation, analysis, indexing, classification and representation of information as well as dissemination, retrieval and use of information – recently mostly in combination with the application of information technology. Results from information science research is therefore also typically directed towards problems in the library domain and used in library science by library professionals. However, recently the growth in development and use of computerised information systems in a much broader spectrum of work domains have resulted in an increased interest and understanding of the importance of this research for design of advanced information systems.

Information science was and has been interdisciplinary in character ever since, drawing on theories and methods from other disciplines such as behavioural science and sociology (user studies and information use), cognitive psychology (cognitive task analysis for system design), computer science (semantic networks), linguistics (indexing, text analysis and representation).

Cognitive Engineering

This approach is concerned with the design of information systems for support of people in their actual work situation based on a systematic analysis of their cogni-

tive tasks and their mental strategies. (Pejtersen 1992a, b, 1993, Rasmussen, Pejtersen and Goodstein 1994; Woods et al. 1988).

The cognitive engineering approach is different from the traditional HCI approach having its focus on the human-work interaction as mediated by a computer rather than on the human-computer interaction.

During a typical work situation, a great variety of information sources are relevant. Observation of the functioning of tools used, readings of measured information, verbal information in case reports and letter files, verbal and non-verbal communication with co-operators, etc. Furthermore, the formats in which the information is retrieved from all these sources should match the needs of a user in a particular work situation. In this situation, the information sources and the "indexing" format will not be standardised and no professional mediator will assist the user. However, in a traditional, "direct contact" work environment evolution and adaptation have granted humans great capability of interpreting perceptively the environment at several levels of interpretation.

When computer based tools are introduced in the interface between people and their work content a very significant information retrieval problem is created. In this situation, computers are mediators in the interaction between the user and the original information sources and a systematic indexing and retrieval strategy must be developed which can match the various information sources to the characteristics of the users and their cognitive tasks. In addition, it must be taken into account, that the users' expertise during a particular situation will vary depending on the uniqueness of the situation and the general competency of the user. Consequently, the retrieval system must be able to respond appropriately to different queries belonging to different modes of interpretation of the information by the user.

Different Representations of Work Situations

In this way, different approaches have been necessary for two "work systems" which represent two extremes of a continuum of work systems stretching *from* the professional user in a well structured work situation, needing access to many different information sources which, more or less, can be designed to match the user needs *to* the situation when a casual user need to access general information services including information sources designed according to a classification of their content without consideration of the actual users needs.

The present design of "integrated work stations" for professional use and the attempts to design "intelligent" information services matching also a casual user's needs call for a more intimate integration of the two approaches. Therefore, several different representations of information are required. Information has to be generated, analysed, formatted and represented to match the information requirements from the different perspectives of the analysis. This calls for a combination of the cognitive engineering design principles and all the well known research theories and methods from information science, as well as advanced information system design concepts together with the skills of professionals from the library science field. The different perspectives of analysis described below were used in the Book House design and lead to the different representations of information as illustrated in figures 7 to 15:

1) A situation independent representation of the subject domain, i.e., the information sources of the work space in which the work will be performed. This is essentially a question of determination of information to be embedded in the database, and hence calls for the well known library and information science expertise of information generation and information management.

2) A structure and representation that matches users' information needs derived from the work requirements facing a user in a particular task situation in terms of goals, constraints and action possibilities. This is essentially a question of development of classification and indexing schemes compatible with categories of user tasks to be used for development of a data base structure. The selection or development of the proper classification, indexing and thesaurus procedures for analysis of information and data base input compatible with users' queries is a well known discipline in information and library science.

3) A representation of various sections of data base content in different formats matching the mental models involved in different retrieval strategies is essential to provide sufficient information in an adequate form to support users' free choice of strategy depending on their subjective preferences and their criteria behind selection of the different strategies. Specific principles for representation of information to special user groups combined with the standardised techniques and practices for information formatting developed for traditional information retrieval systems can be a useful source.

4) A representation of information (i.e. the different sections and formats of information resulted from 3 above) in various interface displays with a form and content that match users' cognitive control of their task depending on their resource profile, i.e. their training and skill and their level of information retrieval expertise.

Very little expertise exists so far in the information and library science field that can contribute fruitfully to this type of interface design – except for the Book House interface design described below.

However, the results of the Book House prototype clearly have shown the successful merging of the cognitive engineering paradigm with the information science paradigm. Many of the features involved in this development can be transferred to "integrated work stations" for other work domains.

The design rationale behind the approach of cognitive system design is described below.

1 The Design Rationale of Cognitive Engineering Design

Human actors are basically goal directed, adaptive organisms. In any task and work environment, the variety of options with respect to "what to do when and how" is great. In general, however, an established practice will evolve within the goals and constraints given by work, guided by the subjective preferences and work styles of the individual actors. Therefore, in order to understand why a particular piece of behaviour is chosen instead of another possible pattern and to be able to predict the response to changes, when new information systems

are introduced, we have to understand how all the action alternatives in a particular situation have been eliminated such that one unique practice can manifest itself. In order to design new information systems, which cope with the end users' task requirements and problem solving activities, we need to know the kind of behaviour to expect, and therefore we have to identify the work depending constraints which define acceptable work performance and the constraints as posed by the tools and means available for work. In addition, we must know the subjective performance criteria which are used by the individual agents to resolve the remaining degrees of freedom. A problem in identifying such behaviour shaping constraints is that *they will not all be active at the time of the behaviour* they control. Behaviour has a prehistory. Patterns of behaviour evolve, they are shaped by prior decisions and choices. When a piece of behaviour is planned by situation and goal analysis and consideration of alternative options for action, the behaviour shaping constraints are being compiled into cue-action patterns and will not be active in later situations, when the particular pattern of behaviour is re-used. It is, however, necessary to identify these "hidden" constraints in order to predict and understand behaviour, even if they are no more needed for control of behaviour. This identification can be difficult, because often they are no more known by the actors and therefore have to be inferred from the work requirements, the resource profile of the actor, and their subjective performance criteria.

In this situation, design of decision support systems cannot only be based on an analysis of the present task performance but must, in addition, focus on creation of a resource envelop around the relevant alternative paths of action. In this way, the individual user will be able to adapt to new requirements and choose a strategy according to the subjective resources and preferences. When design is to be based on the concept of an adequate resource envelope in stead of support of a particular normative work procedure, it follows that separate descriptions are required of: 1. a situation independent representation of the work domain, i.e., the space in which the trajectory of the work will unfold; 2. the work requirements facing a decision maker in a particular situation in terms of goals and constraints and 3. the user's resource profile, depending on training and skill, and subjective preferences which together determine the criteria behind selection of coping strategy; 4. the users' cognitive control of the task depending on their level of expertise.

2 Decision Making in Natural Context

Based on such consideration, we have conducted a series of field studies of performance in actual work environments to have a reliable basis for design of intelligent information and decision support systems. Studies have been made concurrently in libraries, in hospital and manufacturing environments.

From such studies we have identified a couple of characteristics of decision making during work which are important for design of interface systems, and we find it important to distinguish between different modes of decision making in natural work context. First of all, we find it difficult to identify particular, well defined pieces of behaviour as being "decision making." Decision making is not an effort to resolve separate conflicts but has more of the character of a continuous activity to control performance. In other words, decision making is embedded in the more general cognitive control of the state of affairs in the work environment. Secondly, this control depends very much on tacit knowledge, i.e., on intuitive reactions of skilled actors in a familiar context. This latter feature is very dependent on the level of expertise of an actor in the particular situation and it is, therefore, necessary to consider several modes of decision making.

Highly familiar situations. In a very familiar context, behaviour unfolds as an integrated, continuous flow with no need for discrete decisions. Behaviour relies on data-driven chaining of movement patterns generated by the "attuned" internal, dynamic world model. However, conscious mental operations, such as decisions during a preview can play an important role by preconditioning the required dynamic model. Recall in advance of previous, similar situations and rehearsal of the decisions which were necessary, can be required to make the tacit world model sensitive to the cues for future choice.

Ambiguous situations. When this intuitive reaction to events in the environment is no longer effective, mismatch can be experienced between the state of affairs in the environment and the intuitive expectations of the actor. In this case, a number of alternatives for action will normally be perceived by a skilled professional, and the environment will be consulted to read a sign which can resolve the ambiguity. The important point to consider here is that an expert will need no more information than is necessary to resolve the choice between the perceived action alternatives. This mode has several important implications for interface design. One is that decision makers are not subject to "information input" from the environment which has to be analysed, they are asking very specific questions to the environment. They are able to consult "invariants" in the environment by direct perception, if the environment is well structured and transparent.

Problem situations. If no resolution is found in this way, resort will be taken to analytical, symbolic mode of knowledge-based decision making which is the mode which has been subject to most efforts for design of decision support systems.

3 Approaches to Design of Decision Support Systems

Based on the discussion in the previous sections, it will be useful to have a look at the way in which such "natural decision making" can be supported. Normally, decision support system are focused on the *process* of decision making, aiming at advice giving (such as expert system storing the heuristics of expert decision makers), by using computers to do situation analysis and diagnosis by integrating complex data sets by means of normative models of relationships, or by implementing rational planning schemes in computers. In such cases, decision support systems are based on a normative and "rational" conception of decision making. A "decision" is separate in time, information is collected and analysed, goals are evaluated and priorities found, and actions are planned. As it is argued in the previous sections, however, in

natural work situations decision making is a continuous process often shared by several individuals and separate decisions can be difficult to identify.
Therefore, in addition to support of the decision process by advice giving and by supplying an actor with analytic and planning tools, an interface design should be considered for support of natural decision making by presenting for decision makers a complex, rich information context within which they can navigate easily without being constrained by a normative procedure selected by the system designer. We have, elsewhere, called this approach "ecological interface design" (Rasmussen, Pejtersen and Goodstein, 1994).

4 Ecological Interface Design

Designers of complex systems must deal with human perceptual action and cognitive capabilities as they are – this is a given constraint. In other words: we have to observe the laws of nature of the sensor-motor system as well as of the physical environment. Given the remarkable effectiveness of human perception and action in a natural environment, a proper goal would be to design interfaces so as to take advantage of these capabilities. The primary obstacle to overcome is that, unlike the natural environment, in most modern work environments, the goal-relevant system properties are not directly observable unless made so by means of modern information technology. In order to support direct perception, one would have to create some sort of virtual environment that would be as rich in information as the natural ecology. Users should be able to visually explore and manipulate the objects of this artificial environment if direct perception is to be at all possible in addition to the inevitable problem solving activities.
There are certain characteristics of complex work domains that make such an approach a feasible one. Because these are man-made systems, the design is based on intended invariants at various levels of abstraction, which have to be respected during work. In other words, in many cases, there is a lawful basis behind the behaviour shaping constraints, such as basic physical laws for technical systems and policies and regulations for organisations. These invariants provide the basis for the creation of a virtual environment within which actors can navigate. The objective fact of their existence combined with the visualisation power of computer technology provides a new possibility for design of interfaces for direct perception and direct manipulation. The primary goal of ecological interface design is to allow users to take advantage of perception and action, while at the same time providing the necessary support for problem solving activities. In other words, the aim of ecological interface design can be described as trying to make the interface transparent, i.e., to support direct perception at the level of the user's discretionary choice, and to support the level of cognitive control at which the user chooses to perform. To do this, the designer must create a virtual ecological environment, which maps the intended invariants of the functional system design onto the interface. Therefore, a rich support of perception in context is the key design issue.

Book House Design – The Cognitive Basis

In the following, a system for information retrieval and decision support of fiction and fact retrieval in libraries will be presented as an illustrative example of the design principles of a cognitive work analysis discussed above: The design is based on the concept of an adequate resource envelope in stead of support of a particular normative work procedure. Empirical field studies in libraries of the information retrieval task in user-librarian negotiations have been conducted: The cognitive work analysis applied and tested as a method for design and evaluation of the Book House includes (see figure 1 and Pejtersen et al. 1986, 1990, Pejtersen, 1989, 1990, 1991, 1992a, 1992b, 1993):
1. The Domain: Identification of goals, possibilities and constraints: a means ends analysis of the problem space underlying fiction and fact retrieval to make explicit the goals and constraints which must be respected in meaningful work activities, such as what should be done, why should it be done, and when and how can it be done. This analysis defined the constraints for acceptable performance by librarians, who, according to law, shall promote cultural values, not mere entertainment. It also defined the need for embedding information in the Book House system to support these goals.
2. Task Situation: Identification of information sources and use: an analysis of the information retrieval activities and the cognitive decision task, which related to cognitive activities during retrieval. The information gained in this analysis identified the knowledge items in the work domain, which are relevant in the retrieval situation. It also identified the queries which were likely to be made by searchers, when retrieving information. Based on the analysis of these queries, the invariant structure of users' needs and intentions were used to develop a classification scheme for fiction as well as the data base structure and information content of the Book House.
3. Search Strategies: Identification of content of communication in user-system interaction: an analysis of information processing strategies applied by users and librarians during decision making and the underlying mental models.This supplied several coherent sets of mental models, data formats, and tactical rule sets used by searchers with varying expertise and competence. These were used to design a user system dialogue with free access through multiple search strategies and flexible shift among strategies.
4. User Characteristics: Identification of form of interface communication: an analysis of users' cognitive resource profiles necessary to meet the resource requirements of the various strategies. This analysis focused on the level of expertise and the performance criteria of the individual users, which was important to determine the format of interface displays. It was also used to choose icons as interface language as well as a Book House metaphor.The functionality underlying the metaphoric representation was intended to make it possible for users to transfer familiar needs and habits from traditional settings to the new computer environment. Such an approach should be rendered flexible enough to cope with the wide range of library users – novice, casual, professional, children and adults – with more or less analysable needs and intentions.

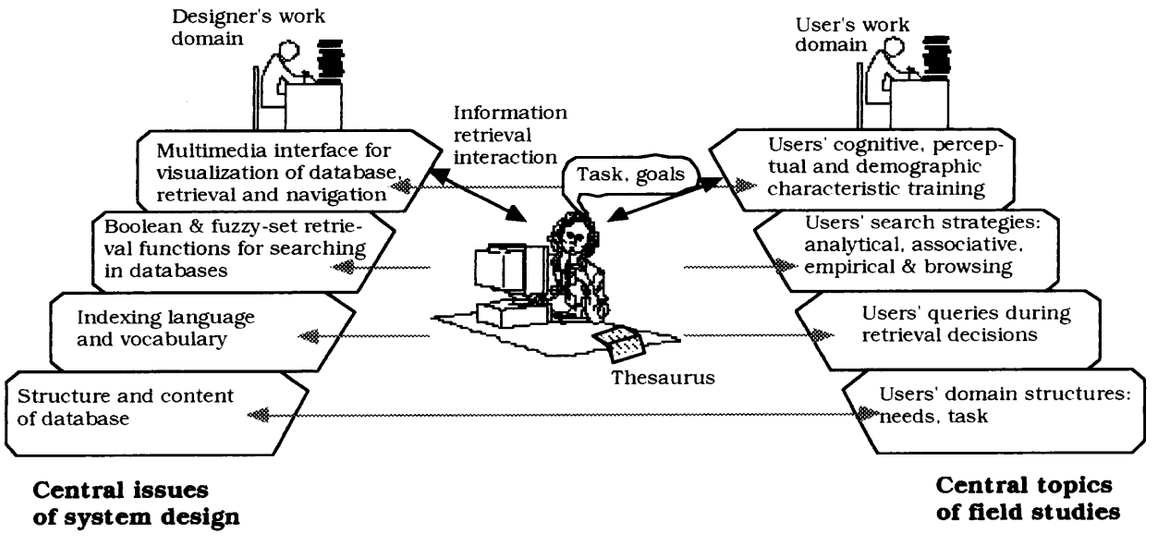

Fig. 1: The figure illustrates the transformations of the key topics of field studies to their representation for system design.

5. Identification of content of communication for co-ordination: an analysis of the allocation of roles among user, librarian and computer tool during the cognitive decision tasks involved in information retrieval. This was used to specify the role allocation among the Book House system and the librarians as well as the subsequent content of communication necessary for the retrieval activity – which led to the design of different types of interfaces to the same database content.

Of particular relevance in the current context are items 1, the analysis of the problem space, 2, the retrieval task analysis for database design and 4, the analysis of user characteristics. We will deal briefly with each of these by describing the results of studies which dealt with the performance of end users and intermediaries carrying out the complex fiction retrieval task in libraries. The studies attempted to uncover their decision strategies and information needs. Verbal protocols were used in a real life setting to gain this knowledge as a basis for the design of new information retrieval tools.

In particular the Book House system is a good example of an ecological interface design, where the invariant structure of the work domain is represented by a metaphor familiar to users, which provides the proper associations for the information retrieval task. It presents for users a coherent display of a complex, rich information context within which they can navigate freely without being constrained by the system. It is intended to support retrieval and decision making in both familiar, ambiguous and problematic situations for casual, novice end users, to whom it provides the capability of visual exploration of icons and direct perception and manipulation of the iconic objects of the artificial, but familiar, environment of the Book House.

The Task Domain

One of the major components of a cognitive task analysis involves the identification of the problem or work space which structures and identifies the information describing the particular “world” under consideration, i.e.fiction – and in a later experiment fact literature retrieval in public library. Today, the retrieval of fiction in libraries is based on an alphabetical classification by authors for shelving arrangements supported by bibliographical databases with author/title access including a number of additional bibliographical data. This system offers the advantage of uniqueness in the arrangement, identification and retrieval of an individual document, but it is of little help for the users with needs related to the subject-matter contents and the other aspects of documents. Requests for specific subjects in fiction like “exciting books about everyday life of children on farms in Guatemala” or “critical books about physical demands in modern sports” are increasing due to the cross disciplinary education in schools and due to TV and other Mass Media. Typically, there are no exact or unique answer to such needs, nor are there within the present computer-based retrieval systems any aid, which the intermediary or end user can utilise to establish a basis for their search strategies. Users face two main difficulties: first the problem of identifying the user’s often subconscious and intuitively formulated need, secondly, the problem of formulating a relevant search strategy among documents that are not indexed and represented according to needs, but according to formal, bibliographical data. What is needed is an appropriate frame of reference for indexing and searching based on a cognitive analysis of the users’ task. This can lead to a system solution which lets the user choose search attributes which adequately cover the specific domain and, at the same time, gives the user the opportunity to solve his/her problem in a natural way.

To support decision making, this work domain should be structured in means-ends relationships, corresponding to a multi-level representation providing descriptions of work substance at different levels of abstraction and aggregation/decomposition. Actually this view of a work domain has broad implications for information retrieval because it implies the need for retrieval of information