A Holistic Profile for Information Filtering Systems

by Kevin R. Parker

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Management Information Systems
at Texas Tech University

May, 1995

ACKNOWLEDGEMENTS

There are many individuals who have played a role in the successful completion of this research. From its inception to its completion, several people have stepped forward to offer their assistance or their ideas.

The extensive literature search could not have been accomplished without the tenacity of Carol Roberts and the staff of the Interlibrary Loan department at Texas Tech University. Susan Dumais and Robert Allen at Bellcore, and Peter Foltz at the University of Pittsburgh, patiently answered my many questions and directed me to others performing research in similar areas.

Each of the committee members provided valuable assistance during the research process. Dr. Ritch Sorenson provided advice on the testing procedure and Dr. Larry Austin contributed to the realization of the conceptual design. Dr. William Marcy provided considerable assistance in prototype development. His prompt answers to my questions and his software expertise helped to reduce the time needed to complete the prototype. Dr. Surya Yadav's guidance, direction, and patience throughout the entire research process made it all possible.

I am indebted to each of the individuals who participated in the comparison study. Their assistance involved a great deal of time and effort, which is greatly appreciated, as is their willingness to get involved in a two week test despite other obligations.

I also wish to thank Jim Phillips, who made his office equipment available to me in order to facilitate the preparation of this document.

To my friends who believed in me, and helped me to believe in myself, I am grateful. David Robinson, Patrick McClure, Toni Logar, Diann Megert, and William McPheron never doubted my ability to endure the trials of graduate school, and their faith emboldened and encouraged me. Thomas Ottaway suffered with me through the long years and was always ready to lend a hand whenever needed. A mere thanks cannot express my gratitude to Tom, but as always he asks for nothing more than friendship.

And last of all, a position to which they have become accustomed during this arduous process, I thank my family. The qualities of perseverance and tenacity were a gift from my parents, as is the realization that there are few limitations for one who is willing to work harder and longer than those around him. I hope that I have made them proud. My wife, who has suffered most of all throughout my time spent in graduate school, has supported me without complaint both financially and emotionally. To her I dedicate this work, and although this is but the first step on a long path filled with such steps, I can at last say, "We did it!"

TABLE OF CONTENTS

ABSTRACT

The quality of information filtering systems is dictated by the quality of the underlying user profile. Most conventional profiles are simply a set of words that describe user interests. While such profiles may be sufficient for casual users, the profile must be much broader in scope before filtering systems prove to be useful in a corporate environment. The creation of existing profiles is unstructured, and may overlook factors that are significant to the decision maker. This paper proposes a holistic profile that addresses this inadequacy by providing a structure to the profile that assimilates not only user interests, but also such considerations as the user's mental model, the user's functional area within the company, concerns related to ongoing projects, client concerns, and contextual information about the organization.

The user's mental model, i.e., his or her internal representation of the how the surrounding world behaves and how to best react to that behavior, is in a state of continuous evolution and can be made richer and more complete by the information gathered by the holistic profile. The holistic profile also enables the decision maker to remain current, informing him or her of opportunities for developing both professional skills and awareness by incorporating features that alert the professional to information that impacts the performance of his or her duties. Ongoing projects can benefit greatly if management or team members are alerted to new developments that might influence the successful completion of the project. Client concerns are a part of the holistic profile, because in order to provide the best client services the user must be aware of external factors that could harm or benefit the client. In addition, because the holistic profile encompasses organizational factors, the information filtering system can serve as an environmental scanning tool for collecting information from the environment to assist in developing strategies that help the organization formulate responses to that environment.

The holistic profile also features domain-specific templates to provide guidance during profile generation, an expansion mechanism to assist in selecting

the most appropriate keywords, and a passive learning mechanism to insure that the holistic profile accurately reflects the user's dynamic information needs.

This research has fully developed the notion of a holistic profile and designed, implemented, and tested a prototype "holistic profile filtering system" to demonstrate the increased effectiveness of the information filtering process. Because the resulting profile embraces the areas of professional development, project management, client services, and environmental scanning, it makes information filtering a more versatile tool with increased usefulness in a corporate environment.

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

Background

Information filtering has recently attracted considerable attention in several professional journals. Despite the apparent advantages of filtering systems, their usefulness is limited by the quality of the user profile. Put more simply, because it is the profile that governs the selection of items, the filtering mechanism is only as good as the profile itself. The scope of existing profiles is often so narrow that the filtering system is severely constrained. Expanding the breadth of the profile to encompass more areas that impact the user enables the filtering system to become a richer and more useful tool.

Information Filtering

The term "information filtering" has been used to describe a variety of processes that involve the selective screening of information, so that a particular user receives only that information in which he or she is interested. In succinct terms, information filtering describes the process of limiting received information by discarding irrelevant information [MORR88]. Information filtering systems monitor streams of incoming data, either broadcast from remote sources such as newswire services, or sent directly by other sources such as electronic mail. The information filtering system filters these incoming streams of information through a user profile and discards data that fail to match that specific profile, leaving only information relevant to the user's expressed interests.

In its simplest form, an information filtering system requires a source of information, a user profile, and a user. The information source provides an incoming electronic stream of information, which is monitored for items of interest. The user profile consists of a set of topics or keywords in which the user has expressed an interest. The incoming information stream is compared to this user profile, and items that do not match the profile are eliminated from the information stream. In this way the system filters out information in which the user is not

interested. After all irrelevant data have been filtered out, the remaining stream, which contains only items that are of interest to the user, is delivered to the user.

Information filtering has been suggested as a corporate tool in many areas, monitoring the information flow "across public and private networks and involving various kinds of information sources, delivery architectures, and user equipment" [LOEB92a, p. 40]. Various applications have been suggested, including filtering technical articles, electronic mail, television and radio broadcasts, and live news services. Additional uses, which are made viable through the use of a holistic profile, will be suggested later in this paper.

User Profiles

The literature contains little information on current methods for creating user profiles. In most cases user profiles are simply described as a model of a user's interests. Foltz and Dumais [FOLT92] note that most often "people provide a set of words to describe their interests" [p.52]. They offer various alternatives, stating that "many other sources of information could be used, such as which articles [users] have read in the past, what organization they work in, or which books they have ordered" [FOLT92, p. 52]. Loeb [LOEB92a, LOEB92b] provides a description of a user profile for the *LyricTime* personalized music system, but few detailed descriptions of more general-purpose user profiles can be found in the literature. Although extensive research has been performed in the area of creating user profiles, such as Stevens [STEV92], Stadnyk and Kass [STAD92], Allen [ALLE90], Teskey [TESK89], Myaeng and Korfhage [MYAE90], Baclace [BACL92], and Goldberg et al. [GOLD92], little work has been published in the area of content or structure of user profiles.

Problem Statement

Current user profiles have been described as a model or a set of words to identify user interests [BELK92, FOLT92]. Interests can be related to domain beliefs, information goals, information types, and information characteristics such as quality and complexity [STAD92]. When limited to personal use such a profile may be

sufficient, yet the potential uses for information filtering go far beyond its use as a personal tool. Conventional user profiles seldom include the appropriate keywords that would permit information filtering to serve as an aid in summarizing information for the user's mental model of external reality, serve as a means of maintaining professional skills and awareness, provide information pertinent to current projects as well as particular client type, and also provide a tool for environmental scanning. The problem stems from the fact that in their current form, most user profiles are relatively incomplete and are unlikely to be sufficiently comprehensive to capture all of the issues that are significant to a decision maker. Nevertheless, the ability to capture these issues is important not only to the individual, but also to the organization. As a part of the organization, the user must be constantly alert to factors that affect the organization and the domain in which the user is operating. Constant monitoring of various information sources for items that impact the individual, his or her professional capacity within the organization, current projects, clients, and the organization's ability to anticipate and react to environmental stimuli, is essential for both the user and the organization.

In addition, information filtering systems seldom provide guidance for the user during the generation of user profiles. With no assistance, the user can only reflect on his or her interests and select the words that best describe those interests based on his or her experience and vocabulary. However, Myaeng and Korfhage [MYAE90] note that people tend to be poor at self-description, so it would seem that some sort of assistance is called for. This research addresses the problem of how to best assist the user in keyword selection.

This research also addresses the problem that users have in selecting the exact words, word combinations, and phrases that are used to describe the information in which they are interested. This problem indicates the need for a method by which the user can expand his or her profile so that the information filtering system will not reject pertinent information simply because the profile and the filtered data use different synonyms to describe that same topic.

Finally, this research must address the issues of changing user information needs and poor keyword selection. The information needs of any particular user are constantly evolving, and as such a mechanism must be provided to allow the profile to adapt to those changes. Not only do some selected keywords become obsolete with the evolution of user interests, but some poorly selected keywords may be seldom used. An adaptive profile is required to address both of these concerns.

Research Issues

User profiles currently serve only as a description of user interests, generally containing a list of words that indicate topics that are of concern to a particular user. In a corporate environment such a profile may be inadequate because additional factors must be taken into consideration. A holistic profile remedies this inadequacy by providing a structure to the profile that incorporates these additional factors. The additional factors included in the holistic profile provide other sources of topics such as the user's functional area within the company, current projects, type of client served by the organization, and contextual information about the organization itself. Such an extension to the user profile helps to insure that not only is the user receiving information that would be provided by conventional user profiles, but is also receiving information that is pertinent to anyone with that particular position, to anyone associated with that project, to anyone providing services to that type of client, and to anyone within that particular industry or organization. These enhancements to the profile result in a holistic profile that is more robust and effective than existing profiles. The term "holistic" refers to the fact that constructing a profile of personal, functional area, current project, client type, and organizational environment segments results in an integrated whole that is greater than the sum of its parts.

An additional research issue that will be addressed by this work is how best to guide a user in the creation of his or her profile. With current information filtering systems, the creation of the user profile is largely a hit-or-miss affair. The user makes a list of keywords that indicate various topics or areas in which he or

she has an interest. This process is highly unstructured, and therefore is likely to overlook important issues that, if they were included in the profile, would benefit the user. Thus, an ancillary research issue that will be addressed is how to provide guidance for the user in generating the profile.

This research will also investigate the issue of how to make the holistic profile as "inclusive" as possible. The concept of *synonymy* refers to the wide diversity in the words that people use to describe the same object or concept [DUMA91, FOLT90]. Various techniques such as thesaurus programs or WordNet have been used to expand the vocabulary and specify various alternatives for the same word or concept. Another issue that will be addressed by this research is how to provide a knowledge base of keyword synonyms or alternatives to make the holistic profile as all-encompassing as possible.

Finally, this research will examine techniques for allowing the profile to adapt to changing user interests. Various techniques such as relevance feedback or frequency of keyword use have been used to address this issue. The final research issue that will be addressed is how to provide an adaptive profile that evolves with user needs.

Research Objectives

This paper addresses the research question of how to augment the user profile in order to unleash the potential of filtering systems. The primary objective of this paper is to conceptually develop and validate a holistic profile to enhance the capabilities of information filtering systems. This research specifies the components that must be included in a holistic profile, provides a structure for developing such a profile, and determines sources of keywords. Specific objectives include not only a determination of issues that influence the specification of user interests in a profile, but also those issues that pertain to the type and source of functional area information, current project information, client type information, and organizational environment information to be included in a holistic profile. In addition, the development of various knowledge-based approaches to guide the user

during profile creation, to expand the profile, and to adapt the profile, were undertaken. In particular, this research had the following objectives.

1. To develop and detail a new concept called a *holistic profile*. This was accomplished by first specifying the requirements for the development of a holistic profile, and then by deriving the structure for holistic profile.
2. To provide a tool to guide the user during the process of profile creation.
3. To provide a technique whereby the holistic profile can be expanded such that it mitigates the problem of synonymy.
4. To provide a technique by which the profile can adapt and continuously evolve.
5. To illustrate the use of the holistic profile in an information filtering application.
6. To provide initial validation for the conceptual development in this dissertation research. The validation of the concepts was achieved through the means of both illustrative and tautological evidence.
7. To provide final validation for the conceptual development in this research through the design, construction, and validation of a prototype holistic profile filtering system.

Research Outcomes

This research investigated the components that must be included in a holistic profile, provided a structured approach for developing such a profile, and determined sources of keywords. It also provided a means of guiding the user through the process of profile construction as well as a means of expanding the profile to make it more complete. The outcomes of the research are:

1. The derivation of and the requirements for the development of a holistic profile;
2. The structure for a holistic profile;
3. A knowledge-based approach to provide assistance and suggestions during the profile generation process;

4. A knowledge-based approach to assist in the expansion of the holistic profile;
5. A passive learning approach to profile adaptation; and
6. A prototype system that demonstrates the feasibility of the holistic profile.

Significance of the Research

The holistic profile is superior to existing profiles because it assists the user in developing a profile that is more comprehensive and better-suited to his or her needs. The holistic profile structures the profile-generating process and guides the user in specifying keywords for inclusion in the profile. Unlike existing profiles, the holistic profile's coverage of topics extends beyond the range of user interests into the domains of functional area, current project, client type, and organizational environment. Thus, the holistic profile is an improvement over existing profiles not only because it provides an expanded perspective, but because its very structure provides guidance to the user in the process of profile development.

The significance of this research stems largely from the effect that the holistic profile has on the capabilities of information filtering systems. With existing user profiles, information filtering is a useful tool, but appears to be one whose capabilities and potential have been untapped. The holistic profile allows information filtering to be used as a tool for assisting the decision maker in using his or her mental model to formulate a decision or strategy. The holistic profile allows information filtering to assist the professional in the areas of growth and development, as well as in the overwhelming task of keeping current with necessary knowledge in a constantly changing professional environment. The holistic profile gathers data that could influence whatever project is currently being undertaken. The holistic profile allows the user to be alert to developments that can affect the client-base. The holistic profile also facilitates the use of information filtering for the strategically significant task of environmental scanning, allowing key players in the organization to become instantly aware of environmental factors that could significantly influence the organization. The holistic profile also provides a structured approach as well as assistance to guide the user in generating the most

comprehensive and pertinent profile possible. The holistic profile is designed to overcome the problem of synonymy, which reduces the effectiveness of many information filtering systems. Finally, the holistic profile includes a passive learning feature to adapt to the user's information needs. Thus, the significance of this research in holistic profiles extends from the individual to the entire organization, making information filtering systems an information systems tool with widespread implications.

Structure of the Dissertation

This dissertation is organized as follows. This initial chapter has provided a statement of the problem that was examined within the context of the proposed research. This chapter also includes a discussion of the background knowledge that has been instrumental in the identification and formulation of this problem. Along with the problem statement is a discussion of some of the issues that are addressed by this research. Those issues were approached by pursuing the research objectives stated in this chapter. The outcomes and significance of the proposed research are also considered.

The following chapter, i.e., Chapter II, reviews current and prior research that is relevant to this dissertation. In particular, literature relevant to information filtering systems, user profiles, mental models, occupational research, project planning and implementation, and environmental scanning is discussed.

Chapter III describes the research methodology that was adhered to in conducting the dissertation research. The unified research methodology was used. A computer-based prototype system serves as a research design to substantiate the conceptual development of this research.

Chapter IV details the development of a conceptual model for the proposed holistic profile. The conceptual model was subsequently used as the basis for a prototype that demonstrated the feasibility of a holistic profile. Based on the knowledge-level and symbol-level principles developed in Chapter IV, the following

chapter (V) discusses the prototype design that guided the prototype implementation, which is the subject of Chapter VI.

Chapter VII describes the validation process for substantiating the conceptual development of this research. The concluding chapter, Chapter VIII, reiterates the contributions of this research in more depth. The key limitations, generalizations, and future work also are addressed in this chapter.

CHAPTER II
REVIEW OF RELEVANT LITERATURE

Introduction

This chapter will discuss some of the theoretical underpinnings that are relevant for this research. This research is a synthesis of several areas of existing work. While research in the areas of information filtering [BACL91, BELK92, DENN82, FISC91, FOLT92, GOLD92, LOEB92a, LOEB92b, SHET94] and user profiles or user models [ALLE90, BACL92, MYAE90, RAM92, STAD92, STEV92, TESK89] serves as the primary basis of this work, other areas upon which this work is based include mental models [CRAI43, JOHN80, JOHN83, KHAZ91, NORM86], occupational research and job descriptions [BENH93, FINE74, GRAN89, LEVI83, MANE88, MAZE91, WOLF88], career mapping [INGR92], occupational advancement [ADAM89, MAGJ93, RAIN93], project administration [PINT90a, PINT90b, LAUF90, SAUN90, DINS90], query expansion [HANC92a, HANC92b, ROBE90], and strategic management techniques such as environmental scanning [AGUI67, ANSO75, DIFF83, ELOF91, FAHE81, HAMB81, HAMB82, HREB85, JENN92, KING87, MCCA92, MESC91, MILL89, MINT76, NARA87, RHYN87, ROBE88, SCHW88, SNYD81, STOF82, STUB82].

This literature review will proceed by examining each of the relevant research areas that provide a basis for the holistic profile. First, the primary research areas that support the concept of the holistic profile--information filtering and user profiles--will be discussed. Next, the research areas that provide the theoretical underpinnings for each of the segments of the holistic profile will be examined. Finally the research areas that support the concept of profile expansion will be considered.

Information Filtering

This research has its primary basis in the area of information filtering. Work in the area of information filtering originated from research into information retrieval [BELK92], which is used to retrieve data from library catalogues and other

databases. The concept of information filtering first appeared in the late 1950's in Luhn's account of the Selective Dissemination of Information (SDI) [LUHN58]. Ackoff [ACKO67] discussed the need for information filtering and condensation in 1967, noting that the two most important functions of information systems are filtration and condensation in order to prevent information overload. Work in the area of information filtering is ongoing at Bellcore, EDS Center for Advanced Research, and Xerox Palo Alto Research Center. The December 1992 issue of Communications of the ACM was dedicated to the topic of information filtering and the related topic of user profiles. Descriptions of various approaches to information filtering can be found in journal articles detailing previous research [ALLE90, BOWE92, FOLT92, GOLD92, LOEB92b, MALO87, STAD91, STAD92, STEV92].

Table 2.1 provides an analysis and comparison of various filtering systems that have been implemented and documented in various journals. These systems encompass a wide variety of information, including music, electronic mail, databases, and newswire services. Encompassing research by Loeb [LOEB92a], Foltz and Dumais [FOLT92], Allen [ALLE90], Malone et al. [MALO87], Goldberg et al. [GOLD92], Bowen et al. [BOWE92], Foltz [FOLT90], Stadnyk and Kass [STAD92], Sheth [SHET94], Baclace [BACL91], and Stevens [STEV92], this comparison clearly shows that these systems are deficient with regard to the user profile. None of the profiles is comprehensive enough to satisfy all of the information needs of a user in an organizational environment, although as Foltz [FOLT90] points out, the user profile must provide "a much larger description of all types of information the user could want" [p.42]. Many of the profiles also lack a provision to deal with synonymy and/or fail to provide guidance for the user during profile construction. Few provide an adaptive profile.

Loeb's [LOEB92a] work on the *LyricTime* music system utilizes a user profile to provide the user with music selections from a database. The user profile in *LyricTime* includes information about the user, time and context information, and user feedback. No mention is made of providing guidance for the user during profile construction and the limited domain precludes the problem of synonymy.

The profile can be adapted by using user feedback to modify the frequency with which a particular type of music is presented.

The work by Foltz and Dumais [FOLT92] compares different methods of matching users' profiles to technical memos and different ways of profiling users' interests. Keyword profiles are used only to provide a comparison to latent semantic indexing. Users were simply asked to compile a list of words and phrases that described their technical interests. No guidance was provided for the user, no provision was made to handle synonymy, and there is no adaptation mechanism. The primary thrust of the research was to test the application of latent semantic indexing to information filtering. Latent semantic indexing handles synonymy by structuring a semantic space that consists of both relevant and non-relevant information, providing examples of both information that a user is interested in and information that fails to meet his or her needs. There is no dependence on exact keyword matching.

Allen [ALLE90] uses the user's past preferences for news stories to construct an implicit user model for retrieving relevant articles. By analyzing what texts were read, as well as the content of the texts, the user model is developed. Because the implicit model is developed automatically, there is no need for user guidance. Synonymy is partially alleviated by using a measure of the overlap of nouns between new and old articles. The ability to allow the profile to adapt is discussed only as an additional research topic.

The Information Lens [MALO87] requires users to construct a rule-based profile to filter mail messages based on keyword matches in the mail fields. Through these rules the user creates a user model to filter information. Tests showed that the majority of the rules were created to match on information about the sender and the recipient. A template editor is provided to assist the user in generating the rules. No provision for synonymy is mentioned and no adaptation mechanism is discussed.

The Tapestry system described in Goldberg [GOLD92] is driven by a set of user queries grouped together to form a "filter query" used to filter electronic mail

or NetNews. The queries resemble database queries rather than keywords. Instead of providing guidance for the user in profile construction, the user is expected to issue a series of individual queries until he or she arrives at the one with the best results. This query is then added to the filter query. There is no provision for synonymy, nor is there any adaptation mechanism aside from user modification.

The Datacycle architecture described by Bowen [BOWE92] monitors the contents of an entire database that is broadcast cyclically. Data retrieval is accomplished through the use of queries similar to conventional database queries. While the system does not include a user profile, it does have provisions for database triggers, which are persistent queries that continuously monitor the database for the appearance of relevant information. The authors note that this feature can be used to implement filtering applications "which involve continuously evaluating information relative to a profile of interest established for individual users or applications" [BOWE92, p. 76]. There is no mention of user guidance. The problem of synonymy is handled somewhat by the use of fuzzy logic, which allows the use of imprecise terms in filtering operations. The adaptability of the "profile" is not provided for beyond that which is accomplished by direct user modification.

In Foltz [FOLT90], the author compares latent semantic indexing and keyword vector matching for filtering NetNews articles. Users rated NetNews articles as either relevant or irrelevant to their interests, and the ratings from the initial 80% of the articles read were used to predict the remaining 20% of the articles. The use of keywords was only briefly discussed, with no mention made of the profile structure or contents, user guidance for keyword selection, or provisions for synonymy with regard to keywords. However, as noted above, latent semantic indexing eliminates concerns about synonymy.

Stadnyk and Kass [STAD92] provide only a brief description of their filtering system, which is used to filter Usenet news messages. The user profile is described as a set of rules correlating conjunctions of description categories with user interest in the message. The categories include domain concepts, goals, message types, message characteristics, and relations. The only form of user guidance discussed is

the use of user stereotypes as initial filters for new users. There is no explicit provision for synonymy, nor is there any mention of adaptation.

The NEWT system developed by Sheth [SHET94] provides a potentially comprehensive profile. The major weakness of this research is that while it attempts to provide for a complete set of user interests, it does so from the perspective of general user interests rather than from the perspective of a corporate user. As in most cases, judicious creation of the profile could potentially remedy this drawback, but is unlikely to occur. A set of agents is provided to assist the user in profile creation. The problem of synonymy is not specifically addressed. The aforementioned agents provide an adaptive profile by utilizing relevance feedback and a genetic algorithm.

Baclace [BACL91] describes an information filtering system that includes a profile that uses document features to create Boolean functions to model user interests. Although user guidance is not specifically provided, it appears to be unnecessary since the profile evolves based on document ratings. There is no specific provision for synonymy. Agents use relevance feedback to adapt the profile based on an economic model used for optimizing computational resources.

Infoscope, discussed in [STEV92], also filters Usenet news messages. The users create matching rules that sort the information into different bins for easier access. The system includes intelligent agents that keep a constantly evolving model of user interests, which is used to make suggestions to the user about the deletion of certain terms from the user model. This not only provides a form of user guidance by allowing the user to edit suggestions rather than remembering each term for manual filter creation, but also provides a form of adaptability. There is no mention of a provision to handle synonymy.

Information Retrieval

As noted above, research in the area of information retrieval provides the foundation for information filtering. "Information retrieval is a well-established field of information science that addresses issues of retrieval from a large collection of

documents in response to user queries" [SHET94, p. 12]. Research into information retrieval encompasses a variety of approaches.

Mauldin [MAUL89, MAUL91] implemented information retrieval with the FERRET system that is based on a text skimming parser. That work had its basis in both information retrieval research and in abstracting systems based on natural language processing [DEJO79, DEJO82]. Jacobs and Rau [JACO90] also used natural language processing to extract information from on-line news services in the SCISOR system. Myaeng and Korfhage [MYAE90], Brooks et al. [BROO85], Belkin [BELK91], and Brajnik et al. [BRAJ87], all consider the application of user modelling on information retrieval. Belkin et al. [BELK93] propose a user interface to support various information seeking strategies. Croft and Turtle [CROF93] also consider retrieval strategies, but with respect to hypertext databases. Chen [CHEN94] examines logic-based information retrieval systems and the use of inference rules. Salton and Buckley [SALT90] consider the use of relevance feedback to improve retrieval performance. Furnas et al. [FURN88], Deerwester et al. [DEER88], and Baker [BAKE62] all study information retrieval based on latent semantic indexing. Palmquist and Balakrishnan [PALM88] look at the use of word association to assist users in query formulation for information retrieval.

User Profiles

Another pivotal research area upon which this research is grounded is user profiles or models. Sherman and Shortliffe [SHER93] note that the idea of user models can be traced back to 1971, when Hansen [HANS71] proposed that the first principle of user engineering is to know the user. Since that time several models of human-computer interaction have been developed.

While much of the literature in the area of information filtering contains references to user profiles with regard to their application to the filtering process, there has been extensive research focusing solely on user models. However, there has been little or no research into expanding the scope of the profile to augment information filtering. Although Allen [ALLE90] examines the use of user models in

the area of information filtering and preference prediction, most research into user models focuses either on permitting interactive systems to communicate with the user in a cooperative manner [KASS89] or on the effectiveness of user models in information retrieval. Kass [KASS89] explores user modeling techniques and how user models can be used in an interactive system. Rich [RICH79, RICH83] looks at developing individualized user models rather than general user models. Teskey [TESK89] examines mathematical representations of user knowledge. Brajnik et al. [BRAJ87] examine the use of user modeling in intelligent information retrieval. Brooks et al. [BROO85] investigate the use of frames to model user interests in information retrieval. Finally, Myaeng and Korfhage [MYAE90] investigate the roles of user profiles in information retrieval.

However worthwhile each of these studies is, they do not permit the full potential of information filtering systems to be realized because they neglect to explore the important issue of extending the scope of topics included in the user profile beyond general user interests. Further, they do not appear to take into consideration the question of how to make information filtering a truly useful corporate tool. As Belkin and Croft [BELK92] noted, researchers "need to do a great deal of research on the dimensions of users' information interests" [p. 37]. This must necessarily encompass the user's professional capacity, the projects with which the user is involved, the types of clients that are being served, and organizational considerations. These concerns must be addressed in order to extend the scope of the profile so that information filtering can become a more effective corporate tool.

The Basis for the Segments of the Holistic Profile

The individual segments of the holistic profile are all based on different areas of research. The theoretical underpinnings for the personal segment, the functional area segment, the current project segment, the organizational environment segment, and the client type segment will be discussed individually.

Personal Segment: Mental Models

The mental model associated with the user plays an important role in the personal segment of the holistic profile. Such considerations allow the personal segment of the profile to extend beyond a general collection of user interests that is common in so many conventional user profiles.

Mental models have their basis in work by Craik [CRAI43]. Johnson-Laird [JOHN80] discusses the development of systems that model cognition. Daniels [DANI86] explores the applicability of cognitive modeling to information retrieval. Belkin [BELK90] considers the cognitive viewpoint in information science. Carlson and Ram [CARL90] examine the use of mental models with respect to hypermedia strategic tools. Yazici and Kluczny [YAZI93] examine the influence of cognitive differences on user interface design and decision making. Brief discussions of mental models can be found in Khazanchi [KHAZ91] and Yadav and Khazanchi [YADA92].

Careful consideration of mental models permits the personal segment of the profile to extend beyond conventional profiles.

Functional Area Segment: Occupational Research, Job Descriptions, Career Mapping, and Occupational Advancement

The topic of employee development contributes to the functional area segment of the holistic profile. It involves a wide spectrum of topics concerning employee skill sets and knowledge and their importance to the organization.

A history of the research areas of occupational research and job descriptions is difficult to trace. Grant [GRAN89] cites the use of job descriptions in the mid 1910's. Manese [MANE88] writes about occupational job evaluation studies. Fine [FINE74] explores functional job analysis studies. Levine et al. [LEVI83] examine job analysis methods. Benham [BENH93] and Rainbird [RAIN93] focus on employee development, while Adams [ADAM89] explores skills development. Ingram [INGR92] looks at career mapping, and Magjuka [MAGJ93] emphasizes employee involvement. Each of these areas merits consideration in the derivation of the

functional area segment, and can potentially contribute to the keyword categories that make up that segment of the holistic profile.

Current Project Segment: Project Planning and Management

Project management and control provide a basis for the current project segment of the profile. Drawing on research in project management, project planning, and project implementation, this area encompasses a great variety of research and research objectives.

Beidleman et al. [BEID90] examine the impact of risk on project success. Dinsmore [DINS90] provides observations about project management from a general business perspective. Saunders [SAUN90] explores project management in the area of research and development. Hall and Hofer [HALL93] investigate the criteria used by venture capitalists when assessing new projects. Laufer [LAUF90] focuses on project planning, while Pinto and Prescott [PINT90b] focus on planning and tactical factors in project planning and implementation. McKim [MCKI90] discusses the essential elements of project control. Rahbar et al. [RAHB91] provide details of a project management knowledge engineering system. Giannotti and Fisher [GIAN93] provide a model for developing a project information system for engineering and construction applications. Pinto [PINT90a] describes the specifics of a project management tool for project tracking and control. Similarly, Howes et al. [HOWE92] describe a project planning and management system for planning and monitoring projects. The primary contribution of Schultz et al. [SCHU87] to this research is a listing of the information needs of project management. Finally, Kern [KERN92] provides a brief discussion of the requirements of project success.

Research into project planning, project management, project implementation, and project control all provide critical insights into the selection of both keywords and keyword categories for the current project segment of the profile. The relevance of any of these areas is dependent on the stage of development of the project in question.

Organizational Environment Segment: Environmental Scanning

The extension of information filtering systems into the area of environmental scanning is made possible by the organizational environment segment of the profile. While the concept of environmental scanning can be traced to Aguilar [AGUI67], Jennings and Lumpkin [JENN92] provide thorough, contemporary coverage of the theoretical background of environmental scanning. Research into strategic management, environmental assessment, and environmental scanning can be found in [DIFF83, ELOF91, FAHE81, MCCA92, STOF82, STUB82, TERR77]. An information systems perspective on environmental scanning appears in Ackoff [ACKO67], Malone et al. [MALO87], Morris [MORR88], and Schwenk [SCHW88]. Ackoff [ACKO67] notes that the two most important functions of information systems are filtration and condensation in order to prevent information overload. Malone et al. [MALO87] investigate the concept of information sharing, or disseminating information only to those who need it. Morris [MORR88] proposes a text-based decision-support system intended to support environmental scanning by actively filtering and condensing text. Schwenk [SCHW88] examines strategic cognition and the ways that environmental factors affect strategic decisions. The capabilities that are made possible by this segment of the profile have been shown by several research areas to be significant.

Client Type Segment: Client Environmental Scanning

The research that supports this segment of the holistic profile is the same research that was discussed in the previous section. In this case, however, the environmental scanning is performed from the perspective of the particular type of client in which the organization specializes. The same principles that make it prudent to monitor the organization environment also indicate that enhanced customer service will result by monitoring the client's environment.

Knowledge-Based Mechanisms

The holistic profile is accompanied by various knowledge-based mechanisms to assist the user in developing the profile. The use of templates to assist the user

during data entry can be seen in such works as Malone et al. [MALO87], and is so widely accepted that no other theoretical basis will be provided. Profile expansion will be provided by a profile expansion mechanism that will rely on some form of synonym knowledge base. This area is well supported by various types of research into profile expansion, as discussed below. The passive learning mechanism will be supported by a profile usage knowledge base in order to provide an adaptive profile. Research into adaptive profiles is detailed last.

Profile Expansion

The purpose of profile expansion is to deal with the problem of synonymy, as discussed previously. Profile expansion has its basis in several areas of research, including query expansion, WordNet, thesaurus programs, and latent semantic indexing. Hancock-Beaulieu [HANC92a, HANC92b], Robertson [ROBE90], and Ekmekcioglu et al. [EKME92] all investigate query expansion from an information retrieval perspective. Miller et al. [MILL90b], Miller [MILL90a], and Gross and Miller [GROS90] explain the use of WordNet for organizing lexical information. Jones [JONE93], Lee et al. [LEE94], Batty [BATT89], Day [DAY92], Brady [BRAD93], Pollard [POLL93], Eastman [EAST88], Wang [WANG85], Wotherspoon [WOTH92], McMath et al. [MCMA89], and Kristensen [KRIS93] explore the use of thesaurus data models in retrieval systems. Furnas [FURN85, FURN88], Foltz [FOLT90], Foltz and Dumais [FOLT92], Dumais [DUMA91], Dumais et al. [DUMA88], and Deerwester et al. [DEER88, DEER90] explore the use of latent semantic indexing to counteract synonymy in information filtering. Each of these approaches provides a viable alternative for profile expansion, and any of them can be adapted to serve as a profile expansion mechanism.

Adaptive Profiles

The literature contains various approaches for providing an adaptive profile. The inclusion of a profile usage knowledge base has its basis in Anderson's Rational Analysis of Human Memory [ANDE90]. He examines several measures of how effectively the history of usage patterns predicts current usage patterns. Anderson's

work also provides the basis for the INFOSCOPE system detailed in Fischer and Stevens [FISC91] and Stevens [STEV92], in which agents are used to make user profiles adaptive. Baclace [BACL92, BACL91] also uses agents to refine the profile based on user feedback and genetic algorithms. Sheth [SHET94] uses both agents and relevance feedback to adapt user profiles. Salton and Buckley [SALT90] explore the use of relevance feedback in information retrieval. This research demonstrates the need for a profile that adapts to the changing information needs of the user.

Summary

The segments of the holistic profile, as well as the use of profile templates, the profile expansion mechanism, and the passive learning mechanism, are well grounded in both information systems research and in other disciplines. Each area that contributes to this research is well supported by the research discussed above. While previous research in the areas of information filtering and user profiles is substantial enough to provide corroboration for this research, the additional studies cited above provide further substantiation for each of the individual areas upon which this research is based.

Table 2.1. Comparison of Filtering Systems.

Comparison Category	Loeb	Foltz	Allen	Malone
Information Type	Multimedia (Music)	Technical Memo Abstracts	News Articles	Mail Messages
Information Source	Music Database	Bellcore Issue	AP Newswire	Email
System Purpose	Entertainment	Professional Awareness	General Awareness	General Awareness
User Type	Casual	Professional (Proactive)	Casual	Casual
Information Lifetime	Indefinite	Limited Duration	Limited Duration	Limited Duration
Delivery Pattern	On demand	Monthly	Continuous	Continuous
Source Availability Pattern	Stored Information	Live Information	Live Information	Live Information
Usage Pattern	Irregulars Intervals	Monthly	Irregular Intervals	Irregular Intervals
Profile Type	Specific--Music Interests	Keywords or Abstracts	User Preferences	Rule-based
Comprehensive Profile	Includes music interests; excludes business interests	Includes list of words describing technical interests; excludes complete business interests	Includes implicit model of user interests only; excludes complete business interests	Includes rules for mail filtering; excludes general user interests
User Guidance in Profile Construction	None Specified	None Specified	Not Needed	Template Editor
Provision for Synonymy	Not Needed	Only by LSI	Limited--Noun Overlap	None Specified
Adaptation Mechanism	Based on user feedback	None provided	Additional research	None provided

Table 2.1. Continued.

Comparison Category	Goldberg	Bowen	Foltz	Stadnyk
Information Type	Mail Messages; News Articles	Database	News Articles	News Articles
Information Source	Email; Newswires; Usenet News	Database Broadcast	Usenet News	Usenet News
System Purpose	General Awareness	Professional Awareness	General Awareness	General Awareness
User Type	Casual	Professional (Proactive)	Casual	Casual
Information Lifetime	Limited Duration	Indefinite	Limited Duration	Limited Duration
Delivery Pattern	Continuous	Continuous	Continuous	Continuous
Source Availability Pattern	Live Information	Stored Information	Live Information	Live Information
Usage Pattern	Irregular Intervals	Irregular Intervals	Irregular Intervals	Irregular Intervals
Profile Type	Filter Query	Database triggers	Not specified	Rule-based
Comprehensive Profile	Includes group of individual queries; excludes general user interests	Includes individual queries; excludes general user interests	Not specified	Includes rules specifying description categories and user interests; excludes general user interests
User Guidance in Profile Construction	None provided	None provided	None provided	User stereotypes for initial filter
Provision for Synonymy	None specified	Limited--Fuzzy Logic	LSI only	None Specified
Adaptation Mechanism	None provided	None provided	None provided	None provided

Table 2.1. Continued.

Comparison Category	Baclace	Sheth	Stevens
Information Type	News Articles	News Articles	News Articles
Information Source	Usenet News	Usenet News	Usenet News
System Purpose	General Awareness	General Awareness	General Awareness
User Type	Casual	Casual	Casual
Information Lifetime	Limited Duration	Limited Duration	Limited Duration
Delivery Pattern	Continuous	Continuous	Continuous
Source Availability Pattern	Live Information	Live Information	Live Information
Usage Pattern	Irregular Intervals	Irregular Intervals	Irregular Intervals
Profile Type	Keywords (document features)	Weighted keywords	Rule-based
Comprehensive Profile	Includes document features to create boolean model; excludes complete business interests	Includes population of profiles to match complete user interests; excludes business model	Includes rules for categorization; excludes general user interests
User Guidance in Profile Construction	Evolves based on document ratings	Agent-based interfaces and pre-created agents	Intelligent agent
Provision for Synonymy	None specified	None specified	None Specified
Adaptation Mechanism	Agents and relevance feedback	Relevance feedback and genetic algorithms	Intelligent agent

CHAPTER III

RESEARCH METHODOLOGY

Introduction

One of the major objectives of this research is the conceptual development of a holistic profile validated by a prototype system. The methodology that is most appropriate for this research is the "unified research methodology" proposed by Baldwin and Yadav [BALD94]. Based on the classical scientific methodology and Ackoff's general research method [ACKO62], the unified research methodology is especially suitable for any research that involves conceptual development that is validated by a prototype. The unified research methodology consists of nine steps:

1. Formulate the problem.
2. Construct knowledge-level principles, concepts, models, or theories that address the problem.
3. Construct symbol-level principles, concepts, models, or theories.
4. Operationalize knowledge-level theories or concepts.
5. Identify or construct a symbol-level design for the system.
6. Identify or develop the prototype system.
7. Test the prototype system.
8. Evaluate and validate the results.
9. Refine the model and repeat steps 1-8 if necessary.

As noted above, this research is conceptual in nature, proposing a structure for an expanded profile that will serve all of the information needs of the user. Step 1 involves the specification of the problem statement that, as discussed previously, concerns the issue of how to make a user profile more complete and sufficiently comprehensive to capture all of the issues that are significant to the decision maker. Steps 2 and 3 comprise the theory building phase of the research. After formulating the problem the knowledge-level concepts and symbol-level concepts must be defined. Step 2 entails the identification of user needs and requirements that dictate system behavior. The goal of step 3 is the architectural design that will best achieve

the desired system behavior, i.e., the constructs used to implement the system. Steps 4 through 8 comprise the theory testing research phase. Theory testing includes the implementation of a prototype system, which embodies the concepts proposed in theory building. Steps 4, 5, and 6 involve the design and implementation of a prototype system. Step 7 involves testing the prototype to determine if it satisfies the requirements specification, and step 8 will evaluate and validate the prototype system by comparing its performance to that of existing profiles. Successful testing and evaluation serve to validate the prototype implementation, which in turn validates the design, which in turn validates the proposed solution. By implication, i.e., transitively, the process serves to validate the overall solution. The final step uses feedback from the previous steps to refine the underlying concepts of the prototype system.

Problem Statement

The problems addressed by this research are threefold. The primary problem is that most conventional user profiles are incomplete and are unlikely to be sufficiently comprehensive to encompass all of the issues that are significant to the decision maker. In addition, there are problems associated with the minimal guidance currently provided to users during the development of their user profile. The final problem is that few conventional user profiles have provisions to account for synonymy, which makes it difficult for users to select the exact words or phrases used to describe the information in which they are interested.

Knowledge-Level Principles for the Holistic Profile

"A way to describe the behavior of systems with wide-ranging capability is in terms of their having knowledge and behaving in light of it" [NEWE90, p. 45]. The knowledge level can best be described in terms of the knowledge required to solve a problem [NEWE82]. The behavior of the knowledge level is based on the principle that a system has a set of goals, and will take whatever actions are necessary to attain those goals given its knowledge [NEWE93]. In other words, the knowledge level refers to the knowledge embodied in the system that allows it to compute the

outputs from the inputs [NEWE93]. Knowledge-level concepts describe how knowledge of the system's environment is used to select the actions that best facilitate reaching the system's goals [NEWE82]. These concepts make it possible to predict and understand the behavior of the system [NEWE82]. Knowledge-level concepts consist of the knowledge required to produce a desired behavior of the system. Levesque [LEVE84] draws an analogy between knowledge at the knowledge level and abstract data types. In order to specify what is required of a desired entity, specify the desired behavior under a set of operations rather than the structures used to realize that behavior [LEVE84].

The knowledge-level principles and concepts that guide the holistic profile research focus on the knowledge required to perform the most effective information filtering for a user, the knowledge that is required to guide the user during profile creation, the knowledge required to overcome the problem of synonymy, and the knowledge required to permit the profile to adapt. The knowledge-level concepts must be checked against the research problem statement for consistency and validity. The following step involves the use of symbol-level concepts to express the method of implementing the stated goals.

Symbol-Level Principles and Concepts for the Holistic Profile

The term "symbol level" refers to the internal architecture that is used to create and represent the knowledge level [NEWE82]. More specifically, symbol-level concepts are representations of knowledge-level objects in terms of symbols that can be manipulated by programs [RICH91]. The symbol-level concepts are the constructs used to implement the knowledge-level concepts. Symbol-level concepts encode knowledge and extract it to make the system behave as desired [NEWE93]. At the symbol level, knowledge corresponds not only to data structures but also to the processes that extract from these structures the knowledge they contain [NEWE82]. Analogous to the preliminary design phase of the software development life cycle, the symbol-level concepts provide a basis for the prototype design specifications.

These specifications include replacing conventional user profiles with a more structured form of profile that encompasses user, functional area, current project, client type, and organizational environment issues. The specifications also call for the provision of support for the user when the profile development process is under way. In addition, provisions to overcome the problem of poor keyword selection must be included. A provision for passive learning must also be incorporated into the system to allow the profile to adapt based on keyword usage. The symbol-level concepts that result from this step must be compared to both the knowledge-level concepts and the problem statement for consistency and validity.

Prototype Development

The next three steps (4-6) of the unified research methodology operationalize the knowledge-level theories in preparation for the design and development of the prototype system. These steps result in a specification for the goals, actions, and control process that were identified in step 2. The internal organization is made explicit in terms of functions, operations, and relationships. The construction of the symbol-level design is similar to the detailed design phase of the software development life cycle. The detailed design will concentrate on enhanced performance as well as user interaction with the system, especially in terms of support and profile expansion features provided by the system.

Once the specifications for construction of the prototype have been established, the prototype can be constructed and tested. A prototype demonstrates "proof of concept" [OLEA88] and will provide an objective demonstration of the concepts that were detailed at the knowledge level. The prototype implementation phase will involve the construction of a holistic profile filtering system. The prototype system will provide the user with a structured approach to generating the profile, as well as providing user support and assistance. It will require the development of a knowledge base of profile "templates," as well as a knowledge base for keyword expansion. Along with a conventional user profile, the resulting

profile will be used in an information filtering system in order to provide a comparison of profiling techniques.

One of the primary strengths of prototyping is that it provides a technique whereby conceptual research can be examined and evaluated. Prototyping can demonstrate or certify that a given model satisfies or fails selected criteria. Prototyping provides a way for designers to assess what has been developed, to determine whether the functionality of that development is useful, and to determine when the conceptual model is complete. Prototypes help to reveal interrelationships, timing, and needed resources. They also provide a more efficient and effective mode of communication. One of the most important strengths of prototyping is that it helps to predict the behavioral characteristics of the entity being modeled. In addition, it makes possible controlled experimentation in situations where direct experiments are impractical or prohibitively expensive.

One of the primary weaknesses is the lack of applicability of prototyping to other research methodologies. In addition, a prototype is by definition not a fully operational system, and thus may not incorporate all aspects of the proposed conceptual model. In addition, there is no single best approach to prototype validation. Prototyping is imprecise, and the degree of this imprecision cannot be measured. Finally, development of a good prototype is often expensive and time consuming.

As the prototype is being constructed, the requirements will be refined and expanded as the need for additional features becomes apparent. Prototype construction necessarily involves some degree of iteration between the previous steps of the development process.

Prototype Testing and Validation

These steps require not only that the prototype system be tested, but also that the functioning model is validated. Testing provides a means of determining if the concepts upon which the prototype is based are both viable and effective. Testing requires the formulation of a set of test cases that are akin to the real data

that the system is intended to manipulate [SOMM89]. Thorough testing should not only attempt to justify that the system functions as intended, but should also attempt to falsify the theories upon which the system is based. Like the classical scientific method, the unified research approach uses the concept of falsification to provide credibility to theories and principles [BALD94]. The most convincing support of a theory arises not from the successful implementation of a functioning system, but from the failure of an honest attempt to falsify that theory [POPP61].

The validation of the prototype system is the subject of a separate chapter (VII) of this document. However, it must be briefly discussed in the context of the unified research methodology.

As with other research methods, it is not enough simply to build the model. Sound scientific practice requires the validation of the methods or techniques by which research is conducted. The goal of prototype validation is to demonstrate or certify that a given model satisfies or fails selected criteria. In this case, that criterion consists of a conventional user profile. Both a conventional user profile and a holistic profile will be developed for the problem domains of public accounting, construction management, and software engineering. If the prototype system meets the criterion of enabling the information filtering system to gather more pertinent data than the conventional profile, and if the process of creating the profile is more intuitive and provides more guidance than in existing cases, then the prototype system will be considered successful. The primary assertion being tested is: There is no difference in the performance of information filtering systems that use conventional user profiles and those that incorporate the holistic profile. If the assertion is rejected, then there is evidence that one type of profile is more effective than the other. Validation is a vital step in the research process because it helps to establish the acceptability of the prototype's results, and thus generates confidence that the prototype system's performance accurately reflects the conceptual system being modeled.

Refinement

Unsatisfactory or unexpected results discovered during testing and validation may require some modifications in the system design. In such cases it may be necessary to reiterate steps 1 through 8, correcting deficiencies or augmenting the design.

Summary

A great deal of research has relied on prototype development to demonstrate or prove various theories. However, unless such prototypes are part of an overall research methodology, the results of such works are open to questions and doubt. The unified research methodology provides credibility to the prototyping process by organizing the research process into problem statement, theory building, and theory testing phases.

CHAPTER IV

CONCEPTUAL DEVELOPMENT

Introduction

The first chapter of this document provided the background for this research, and then specified the problem statement, research issues, research objectives, and research outcomes. The preceding chapter explained the research methodology that is used and briefly discussed the requirements of the system. This chapter will review those requirements and then will detail the conceptual development of the solution that serves as the basis of this research.

System Behavior

Without information pertaining to mental models, professional development, project management, client support, and environmental scanning, today's decision makers are not as effective as they otherwise could be. An effective information filtering system should support the user in addressing these concerns. However, the information filtering system is dependent on the user profile for the variety of information that is amassed for the user. Thus, improving the breadth of information yielded by the information filtering system requires improvements in the profile. Further, the profile should provide the user with some type of structured approach or guidelines for compiling the profile. In order for these goals to be achieved, user profiles must not only be given a structure, but must also be expanded into a comprehensive or holistic profile that addresses all of these concerns.

The first concern that must be addressed by information filtering systems is that of the mental model. A mental model is a small-scale model of reality that each person has formed based on his or her experiences and biases. It influences the way in which a decision maker views and reacts to various situations. The profile should incorporate features to help consolidate information for the decision maker's mental model in order to assist him or her in making decisions.

The information filtering system must also assist the user in the areas of professional growth and development. A decision maker must be committed to staying on top of his or her field by keeping abreast of opportunities for developing his or her professional skills and awareness. Profiles should be designed to incorporate features that alert the professional to new developments, whether educational, technical, or regulatory, that impact the performance of his or her duties. The professional should also be aware of information that will help attain the additional skills and knowledge necessary for advancement within the organization.

The information filtering system should provide assistance in the area of project administration. Any information that might affect the planning or implementation of current projects should be made available to the user. Only with current information can project administration be successful, and filtering systems can help to provide that information.

Perhaps the most important potential use of information filtering systems that is seldom exploited is the capacity for environmental scanning. "Organizations are facing increasingly complex environments in terms of the number of coexisting issues competing for attention and resources" [MCCA92, p. 394]. This mass of ideas, events, and concepts in the organization's environment poses significant strategic implications for decision makers [MCCA92]. Profiles should be extended to include organizational factors. This will allow information filtering systems to act as an environmental scanning tool for collecting information from the environment and for assisting in developing strategies that help an organization formulate responses to that environment. Such information as it pertains to the clients served by the organization is also essential.

Additionally, most users would benefit from some form of guidance in the compilation of their profile. Without such direction it is likely that the user will overlook several significant keywords that should be included in the profile. The profile should be structured to facilitate the profile creation process by providing domain-specific suggestions.

The system should also help the user in compiling the most thorough profile possible. Because it is difficult for users to arrive at the exact keywords that will provide the information that is needed, the system should supplement the keyword selection with domain-specific synonyms to make the filtering process more successful.

Additionally, the profile should be capable of adapting to the user's changing information needs. Fischer and Stevens [FISC91] note that such adaptation must be based on an analysis of user behavior patterns of which users are not aware. The system can assist users in analyzing this behavior in order to modify their profile. This feature also helps to increase the precision of the filtering system by allowing the removal of imprecise keywords.

Based on these considerations the system behavior can be identified. Only if the system is capable of performing these functions can the system in question achieve its overall goal of providing the capabilities that the user requires. The system behavior that will permit the holistic profile to address all of the issues that are enumerated in the problem statement includes the following:

1. The level of recall achieved by the information filtering system should be improved. Recall refers to the number of relevant documents retained by the filter [MAUL91].
2. The comprehensiveness of the information retained by the information filtering system should be enhanced. The profile should reflect the entire spectrum of personal and organizational concerns.
3. The filtering system should be an integral part of building the organization's information-based comparative advantage. This refers to the development of a relative advantage in the competitive marketplace on the basis of superior information or knowledge [KING87], which can be obtained through environmental scanning.
4. The information filtering system should include a mechanism or feature to guide the user in the creation of a profile. Because conventional user profiles

are simply collections of user interests [FOLT92], the profile generation process currently is unstructured and largely hit or miss.

5. The information filtering system should provide an expansion mechanism to supplement the user's keyword selection. Users seldom select the keyword that results in the best payoff, so some method of suggesting supplemental keywords is needed [FURN83].
6. The information filtering system should provide a profile that adapts to the user's information usage patterns. One of the major problems associated with the use of information filtering is the effort required to maintain and evolve the profile over time [STEV92], so some form of adaptive profile is required to assist the user in this task.
7. The information filtering system should provide a profile that adapts to user feedback regarding the precision of the holistic profile. Precision refers to the relevance or accuracy of the retained documents [MAUL91].
8. The holistic profile should be segmented so that domain-specific segments, such as functional area or organizational environment, can be duplicated and shared with other members of the functional area or organization.

Knowledge-Level Concepts and Principles

Once the system behavior has been determined, the conceptual development progresses by determining the knowledge required to make that behavior possible. This conceptualization of the knowledge level begins with the derivation of knowledge-level concepts or theories.

Knowledge-Level Concepts

Knowledge-level concepts are based on the idea that the system must maintain knowledge of certain things in order to behave as required. In other words, the knowledge level embodies the knowledge required to produce a desired behavior of the system. The knowledge-level concepts include the holistic profile, domain-specific templates, domain-specific synonyms, and keyword-performance.

Each of the knowledge-level concepts below will be accompanied by a brief explanation.

Holistic Profile

Because the performance of the information filtering system is so heavily dependent on the quality of the profile, the user profile must reflect the user's interests across the entire spectrum of both personal as well as organizational considerations. Unless these elements are taken into account the profile--and by extension the filtering system--will be unable to provide the full range of data gathering services that would otherwise be possible. A holistic profile will provide a structure for the user model upon which the filtering process is based. This holistic profile will be made up of distinct segments consisting of keyword categories that fully describe the user's interests. In this way not only will the filtering system gather information that satisfies personal interests, but it will also gather information relevant to occupational, ongoing project, organizational, and client considerations as well, making the information filtering system as robust as possible.

Domain-Specific Templates

Because a user is seldom able to formulate an entire set of appropriate keywords to describe a particular domain of interests, the system should provide the user with a set of templates that can provide guidance. This requires the profile creation mechanism to include a knowledge base of templates made up of domain-specific keywords associated with a template key. When the user enters a template key, i.e., a domain-specific keyword for which a representative set of interests exists, e.g. *public accounting* or *software engineer*, the profile creation mechanism must be capable of bringing up a template consisting of keywords that can either be used "as is" or that may direct the user's attention toward a subject area that may have otherwise been overlooked.

Domain-Specific Synonyms

Users seldom select the keyword or phrase that best describes their information interests. Along the same lines, although the user may select a suitable keyword, the information provider may have selected a completely different keyword with the same meaning to describe the subject area, resulting in a filtering "miss." These problems can be alleviated if the system has knowledge of a set of domain-specific synonyms.

Alternative or supplemental keywords will be suggested to the user by means of a knowledge base of keyword synonyms. When the user indicates a keyword for which a more appropriate keyword or alternative keywords exist, the system will display the alternatives and allow the user to supplement his or her choice. Indicating related keywords to the user results in a profile that is less susceptible to the problems of synonymy, and assists the user in selecting a wider range of pertinent keywords to describe his or her interests.

Keyword-Performance

In order to maintain an accurate representation of the user's information interests and to maintain a profile made up of productive keywords only, the profile must be dynamic and capable of adapting. The information needs of the user evolve over time and the holistic profile must be capable of keeping pace with these changing interests. In addition, some keywords may be poor choices in that they result primarily in the retention of documents that are of no interest to the user and therefore reduce precision (quality). By maintaining a knowledge base of keyword usage patterns and soliciting user feedback, it is possible to assist the user in making alterations to the profile so that there is an accurate correlation to the user's information needs. If a keyword is used infrequently or results in the retention of irrelevant documents, then it will be called to the user's attention for possible removal or elaboration.

Knowledge-Level Principles

The next step in the conceptual development is to derive a set of knowledge-level principles. These principles are propositions that describe the relationship between system behavior (B), the environment or domain (E), the symbol level (S), and knowledge-level concepts (K).

1. If the information filtering system has knowledge of the holistic profile (K), then the system will retain a greater number of documents that relate to user interests (B). That is, the level of recall will improve.
2. If the information filtering system can retain knowledge of personal interests (K) in one module of the holistic profile (S), then the system will be able to provide the user not only with information that satisfies his or her general interests, but also with input for the user's mental model (B).
3. If the information filtering system can retain knowledge of functional area interests (K) in one module of the holistic profile (S), then the system will be able to retain information to assist the user in increasing his or her professional knowledge and skills in order to remain current in the field and to qualify for advancement (B).
4. If the information filtering system can retain knowledge about ongoing projects (K) in one module of the holistic profile (S), then the system will be able to gather information vital to the outcome of those projects (B).
5. If the information filtering system can retain knowledge about the organizational environment and client environment (K) in individual modules of the profile (S), then the system will contribute to the organization's information-based comparative advantage by providing an environmental scanning tool--not only for corporate purposes--but also for client support (B).
6. If the holistic profile development system has knowledge (K) of domain-specific templates and keywords (E), then the system will be able to guide the user in the development of the profile and also assist in overcoming the problem of synonymy by providing alternative keywords (B).

7. If the information filtering system uses a performance object (S) to retain knowledge of both the keywords responsible for retaining documents and of the user's perception of the relevance of those documents (K), then the system can evaluate the usefulness of the keywords in the holistic profile and suggest removal of keywords with a history of poor performance (B).

Symbol-Level Concepts and Principles

As described in the previous chapter, the symbol level refers to the internal architecture that is used to represent the knowledge level. The symbol-level representation encodes knowledge and extracts it to make the system behave as desired [NEWE82]. At this stage in the conceptual development various architectures are evaluated and different knowledge-representation techniques are analyzed to determine which one best supports the knowledge-level concepts.

Symbol-Level Concepts

The first step in conceptualizing the symbol level is to formulate a set of symbol-level concepts. These concepts focus on the architecture necessary to encode the knowledge required by the system. The symbol-level constructs include modularized architecture that can be expressed through an object-oriented representation.

Modularized Architecture

A modularized architecture will be used to represent the diverse areas of user interests, including personal, functional area, ongoing project, organizational environment, and client type concerns. The modularized structure allows the addition of supplemental modules representing specialized areas of user interests and therefore insures a comprehensive model of user interests.

The modularized architecture also makes it possible for individual segments to be replicated for use by other users in the same department or organization. Further, the modularized structure makes it possible for the system to provide

domain-specific templates for specific segments to assist the user in specifying his or her interests.

Object-Oriented Representation

There is a variety of ways in which the modularized structure can be represented. Techniques such as frame-based, object-oriented, formal logic, and neural network representations are all suitable [DIET93]. An object-oriented design was selected to represent the modularized structure.

An object-oriented approach makes it possible to use several object types to implement the various components of the modularized structure, as well as several other specialized features. For example, the modularized structure can best be represented as a set of objects that embody the necessary keyword categories that make up each of the segments of the holistic profile. These segment objects can be used to encode the personal interests, functional area interests, ongoing project interests, organizational environment interests, and client type interests. Each of these segment objects is made up of keyword-category objects that represent groupings of user interests. These objects are composed of keyword objects that represent user interests.

Precision Objects

Each keyword object has associated with it a precision object. A precision object consists of attributes that record such details as frequency of use and performance rating.

Synonym Objects

A synonym object includes an attribute representing the keyword and multiple attributes representing potential domain-specific synonyms.

Symbol-Level Principles

Once the symbol-level concepts have been determined, the conceptual development proceeds by relating those concepts to the system behavior and to the knowledge that makes that behavior possible. The symbol-level principles express

the relationships between the symbol level, the system behavior, the knowledge level, and the environment.

1. It must be possible to create a holistic profile of user interests that will insure greatly increased levels of both recall and comprehensiveness of retained information. Such a profile will require that all pertinent areas of user interests are identified and accounted for, and that the majority of the features of each of those areas can be grouped into a set of keyword categories.
2. It must be possible to evaluate the degree of success with which each keyword results in information retention. A structure that records frequency of use as well as relative success can sufficiently describe keyword performance and system precision. Such a structure can suggest profile modifications and provide a pseudo-adaptive quality.
3. A modularized structure allows the isolation of domain-dependent segments. This allows the system to include a guidance feature to assist in the specification of those segments. These domain-specific templates can be developed only with the assistance of a domain expert.
4. A modularized structure allows the isolation of segments that may be domain dependent. This allows those segments to be duplicated and shared by other users in the same domain.
5. A structure that stores domain-specific keyword synonyms assists the user in the specification of keywords. A domain-specific knowledge base of synonyms can be developed only with the assistance of a domain expert.

Strategy for Conceptual Development

Because of the wide range of system requirements, the strategy for the conceptual development of the holistic profile followed a four-stage approach.

Stage 1: Specify a structure for the holistic profile that will satisfy the information needs of the user. This stage satisfies the first five knowledge-level concepts.

Stage 2: Develop a knowledge base of sample profiles, i.e., specialized templates, to suggest keywords to the user during the profile generation phase. This partially satisfies the sixth knowledge-level concept.

Stage 3: Develop a knowledge base of keyword synonyms so that the user's initial profile can be expanded in order to alleviate the problem of synonymy. This helps to satisfy the sixth knowledge-level concept.

Stage 4: Develop a passive learning system to allow adaptation of the profile based on usage patterns. This satisfies the seventh knowledge-level concept.

System Architecture

The system architecture reflects the four stages noted above. In this architecture the profile creation mechanism is utilized to assist in the creation of the profile. As the user inputs his or her interests, the profile template knowledge base provides templates containing keyword recommendations. The result of this phase is the unenhanced holistic profile. This is passed to the profile expansion mechanism, which indicates to the synonym knowledge base the keyword to be expanded, and is provided with keyword synonyms. The outcome of this process is the enhanced holistic profile, which is modified by the passive learning mechanism based on usage patterns of the various keywords. The final outcome of the overall process is the adaptive holistic profile. The system architecture is shown in Figure 4.1.

Stage One – The Holistic Profile

The proposed solution for stage one is a holistic profile that incorporates the following segments:

(1) a personal profile containing the interests of the particular user;

(2) a functional area profile, containing those topics that should be of interest to any employee holding that particular position;

(3) a current project profile, containing information topics that may affect the project planning and implementation activities of any ventures currently being undertaken;

(4) an organizational environment profile, made up of topics that influence any employee associated with that particular organization; and

(5) a client type profile, made up of topics that influence the operations of any client organizations.

The structure of the holistic profile is shown in Figure 4.2.

Personal Profile

The personal profile closely resembles existing user profiles, consisting of a set of topics or keywords in which the user has expressed an interest. This set of topics may include items of general interest, items pertaining to the user's area of expertise, or areas in which the user wishes to enhance his or her knowledge and understanding. The personal profile normally contains topics of specific interest to that particular user.

One consideration that must be taken into account when constructing the personal profile is the individual's mental model. Before explaining the relationship between the personal profile and the mental model, it is necessary to first establish what is meant by the term "mental model." A mental model is an internal representation of reality formed from the individual's experiences and concept of reality. Norman [NORM86] explains that the concept of mental models is based on the notion "that people form internal, mental models of themselves and of the things and people with whom they interact. These models provide predictive and explanatory power for understanding the interaction" [p. 46]. He also notes that while the models are neither complete nor accurate, they nevertheless function to guide much human behavior [NORM86]. They guide human behavior by enabling "individuals to make inferences and predictions, to understand phenomena, to decide what action to take and control its execution, and above all to experience events by proxy" [JOHN83, p. 397].

The personal profile can contribute to the shaping of the user's mental model. "Mental models evolve naturally through interaction with the world..." [NORM86, p. 46]. The nature of the interaction, combined with the person's prior knowledge and understanding, help to guide that evolution [NORM86]. Because the profile can enhance not only the user's prior knowledge, but also the interaction itself, filtering systems make the mental model more effective.

The concept of a mental model is consistent with the definition of thinking as the manipulation of internal representations or models of reality.

> If the organism carries a "small-scale model" of external reality and of its possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and the future, and in every way to react in a much fuller, safer, and more complete manner to the emergencies which face it. [CRAI43] (cited in [JOHN80, p. 73])

These mental models provide a perspective that helps the decision maker to diagnose, formulate, and/or understand problems from scrutiny of the environment [KHAZ91]. Mintzberg notes that

> In effect, managers (like everyone else) use their information to build mental "models" of their world, which are implicit synthesized apprehensions of how their organizations and environments function. Then, whenever an action is contemplated, the manager can simulate the outcome using his implicit models. There can be little doubt that this kind of activity goes on all the time in the world of management. [MINT76, p. 54]

While various tools such as executive information systems (EIS) can be used to support the information requirements of management by scanning environmental information and summarizing it for the executive's mental model of the business, an information filtering system with an appropriate personal profile can supplement the process. While an EIS can provide a manager with some of the information required for his or her mental model, filtering systems permit the filtering of additional sources of information and thus make the mental model more effective [CARL90].

Consequently, the generation of the personal profile must take into account not only general user interests, but also how information is influenced by factors such as how the profile can be used to contribute to the user's mental model.

Functional Area Profile

Functional area information is necessary because certain topics are relevant to a specific job or position but may be overlooked when generating the user profile layer. Each functional area has associated with it certain areas of knowledge with which the user must be familiar. This information may be of a more specific nature or of a narrower scope than that which is requested in the user profile. For example, an accountant specializing in estates may have a user profile that includes such topics as estate planning, but the functional area profile would include more specific topics such as estate and gift tax. The functional area profile provides a safety net for the user, specifying specific categories of information that are relevant to the position, but which are taken for granted and often overlooked when specifying the user profile. A functional area profile can help to guide the user toward interests associated with his or her position or functional area rather than toward the general interests specified in the personal profile.

The functional area profile should be made up of topics that are required knowledge for a particular functional area. One potential source of keywords for the functional area profile information is the job description. A well-written job description will include knowledge priority, areas of expertise, and knowledge requirements [WOLF88]. If the job analysis process was properly conducted when formulating job descriptions, sufficient information about each position will be available to provide specifics for a functional area profile. As a potential source of functional area data, a well-written job description will include a work setting portion that describes the context or conditions under which the work is performed as well as a list of task statements organized into major duties that embody the content of the job [MANE88].

Job descriptions are sometimes accompanied by, or even incorporate, a job specification. A job specification details experience, level of education, and the kinds and levels of skills, abilities, knowledge, and so on that one must possess in order to perform a job well or up to standard. The job specification details the kinds and amounts of experience and training needed by employees if they are to acquire the skills, knowledge, and abilities needed to satisfactorily perform the job [GRAN89]. Wolff [WOLF88] refers to this as the Skills/Characteristics section, and includes knowledge priority, areas of expertise, knowledge requirements, and abilities. Such items in the job specification can serve as an excellent source of keywords for the functional area profile.

Another feature that is often incorporated into the job description is a description of the skills, knowledge, and abilities at which employee should become proficient while serving in that position. This section is included to show employees what additional skills and knowledge that they need to acquire in order to advance in the company. This career progression section is another potentially valuable source of functional area profile keywords.

If current job descriptions are not sufficient, there are more general sources of functional area information such as occupational guides. One example is The Enhanced Guide for Occupational Exploration, which provides such detailed descriptions of various occupations that it can serve as a general guideline for the selection of keywords related to that particular job [MAZE91]. Figure 4.3 shows an example of an occupational description. Note that while some keywords are evident, others are suggested only indirectly.

Current Project Profile

A project can be defined as "a subset of a company's activity, a finite job, with its own objectives and budget" [SAUN90, p. 45]. Several project management and control tools are available [RAHB91, GIAN93, HOWE92, PINT90a, MCKI90], but most are concerned with critical successful factors and details about the status of the project itself. Few, if any, are concerned with gathering external information

despite the fact that external factors often play a significant role in the successful completion of a project. Rahbar et al. [RAHB91] point out that such data gathering, which they refer to as influence analysis, is necessary because events or conditions external to the project and outside of the control of the project manager may influence the successful completion of the project. Schultz et al. [SCHU87] also note the importance of external factors by pointing out that

> The strategic factors in project management tend to require a wider (more diverse and external) collection of information. For example, it is important to monitor the activities of competitors, future technologies, social and political changes, and so forth. [SCHU87, p.38]

Research in the areas of project management, project planning, project implementation, project control, and project information requirements provides an indication of the types of information that are vital to project success. Sources of keywords to describe critical factors in the planning and implementation of projects can be found in many areas. For instance, "project objectives include scope, time and cost parameters, priorities, environmental, technological, and operational constraints and restraints, milestones, and control considerations" [LAUF90, p. 168]. Project management typically includes programming, scheduling, cost control, quality control, scope management, contract management, resource management, and interface management [DINS90]. Project attributes such as market considerations, product differentiation, economic environment, material costs are also important factors to consider [HALL93]. Project stakeholders, i.e., those who have a stake in project outcome such as clients, users, managers, financiers, suppliers of technology, and higher management, must also be considered [DINS90]. Factors such as technical requirements and manpower capabilities provide additional sources of information that are vital to project control. For example, if a project involves the purchase of one hundred microcomputers for the office staff, developments such as the discovery of flaws in the Intel Pentium chip, limited production of clock-tripled chips, or widespread acceptance of the PowerPC chip could have an effect on the scheduling of the purchase. The financial requirements

of a project are also influenced by external information [HALL93]. The recent collapse of the Bell Atlantic-TCI merger was attributed by many industry observers to FCC regulation of cable rates. Additional factors such as return on investment may also be affected by external factors.

Because the current project profile is concerned primarily with ongoing projects, factors stemming from project implementation make up a large part of the profile. Project implementation is a two-stage process, consisting of initial planning and subsequent tactical operationalization [SCHU87]. Planning includes the critical success factors of project mission, top management support, scheduling, planning, and client consultation [PINT90b]. Tactical operationalization includes personnel, technical tasks, client acceptance, monitoring and feedback, communication, and trouble-shooting [PINT90b].

Projects in both the planning and the implementation phases benefit greatly from immediate knowledge of environmental factors that may affect project quality as well as the project's successful completion. The type of information needed is dependent not only on the phase that the project is in but also on the overall goal of the project itself. Effective project implementation requires that those associated with the project gather information in order to reexamine both the feasibility and the appropriateness of project objectives and planning assumptions on an ongoing basis [LAUF90]. The need for current information in the project planning and implementation phases is reinforced by the fact that only around twenty percent of the projects that are undertaken are successfully completed [BEID90]. Some of the causes for this low success rate include technical failure, poor management, and legislative or regulatory changes. One key to successful project administration is to identify and manage these risks by being made aware of pertinent information as soon as it becomes available [BEID90].

Organizational Environment Profile

Organizational environment information is necessary because certain topics are relevant to organizations and their specific context and may not be apparent

when formulating other segments of the profile. Each organization has associated with it a particular context pertaining to customer attitudes, competitors' actions, regulatory patterns, technological trends, and so forth. This information, which pertains specifically to the type of organization by which the user is employed, may be overlooked or omitted from the other segments of the profile. For example, if an insurance company writes the bulk of its homeowners policies in the Miami area, the organizational environment layer of the profile may include "Miami" as a keyword, alerting users on an organization-wide basis of unusual circumstances in the Miami area that might result in excessive claims. Such a keyword may have been overlooked in the other segments of an insurance adjuster's profile.

A well-designed organizational environment profile can enable an organization to achieve an information-based comparative business advantage. This refers to the development of a relative advantage in the competitive marketplace on the basis of superior information or knowledge [KING87]. One means by which an information-based strategic comparative advantage may be achieved is by including an organizational environment profile in the filtering mechanism, which enables that filtering mechanism to be utilized as an environmental scanning tool. Environmental scanning was defined by Aguilar [AGUI67] in 1967 as the process of gathering information about events and relationships in the organization's environment, the knowledge of which assists in planning future courses of action. It entails perceiving and interpreting both the internal and external environment with the objective of making appropriate operational, tactical, and strategic decisions that help insure the success of the firm [ELOF91]. Any organization that fails to continuously examine the environment in order to determine the conditions under which it must operate courts disaster [MITR85]. Identification of key economic, social, and technological issues that affect the organization, its life cycle stages, and their relevance to each other helps managers to allocate attention and resources to them [MCCA92]. Scanning is a fundamental, early step in the chain of perceptions and actions that permit an organization to adapt to its environment [HAMB81].

Aguilar [AGUI67] stresses the close relationship between strategic planning and scanning, noting that scanning is the acquisition of external strategic information that is useful for making decisions about company strategy and long term plans. Jennings and Lumpkin [JENN92] report that the objectives of environmental scanning vary with the business strategy employed by an organization. Differentiation strategy is associated with a systematic scanning activity to alert the organization to market opportunities as well as indications of innovations [MILL89]. Cost leadership strategy involves scanning for more efficient methods of production as well as innovations made by the competition [MILL89]. Reactive strategy is associated with scanning the external environment for problems [ANSO75], while low cost strategy directs the scanning effort toward solving specific problems regarding product cost [HREB85]. An organization's strategy determines whether environmental scanning is used to search for opportunities or to forewarn of threats [SNYD81]. The goals of an organization are continuously evolving, and as they are changing, so too are the pertinent threats and opportunities that must be monitored. As the goals of the organization change, new threats and opportunities must be monitored [ELOF91].

An increasing number of potential sources for environmental information are becoming available. Morris notes that

> If an information system could "pre-scan" the potential sources...then managers would be able to reduce the time spent in scanning activities, increase the number of information sources covered, and better focus their scanning efforts. [MORR88, p. 8]

Any environmental scanning system is dependent on the identification of pertinent factors to be scanned. The organizational environment profile should be composed of factors that are likely to have a major impact on the firm. Based on the current goals and strategic approach of the organization, qualitative indicators that provide insight into threats and opportunities must be identified and evaluated [ELOF91]. Rhyne notes that

> Within a given industry, there are usually a few key factors on which all firms depend for success. These may be a widespread level of

> manufacturing technology, certain characteristics of customers, the availability of low-cost raw materials, or the presence of protective regulation. Changes in these factors may dramatically affect the industry as a whole or may provide a major competitive advantage if an individual firm achieves a breakthrough. For example, if all firms are using the same basic technology, a new method of production can give one competitor or a totally new firm a competitive advantage while rendering other firms' production obsolete. [RHYN87, p. 401]

It is essential to recognize that selecting topics or keywords for inclusion in the organizational environment segment of the profile is an extremely ill-structured and ambiguous activity. There are essentially no limits to the potentially relevant topics. The categories of data are inherently scattered, vague, and imprecise. Thus, the purview of the keywords must be broad, but there are no guidelines as to where the profile should be focused. As a result, it should be expected that the filtering process will allow a rather large degree of insignificant information to pass successfully through the organizational environment segment of the filtering profile [NARA87].

In developing the organizational environment profile, it is necessary to identify environmental factors, both external and internal, that can impact the organization's future. The internal environment includes products, geographic markets, and user segments. The external environment includes competition, external threats or opportunities, and strategic vulnerability areas such as raw materials, technology, labor, legislation, and capital [ROBE88].

Some general factors that may be considered when forming the organization's profile include:

1. Customer base and marketplace.
2. Industry type.
3. Socioeconomic.
4. Demographic.
5. Competition.
6. Corporate environment.

7. Manpower and worker skills.
8. Raw materials and capital resources.
9. Technological and scientific (general and specific products).
10. Geopolitical.
11. Government and politics.
12. Business law.
13. Patent and trademark law.
14. Taxes.
15. Import/export law.

Many authors have argued that the business environment is changing more rapidly and less predictably than in the past, making it more difficult to anticipate that change. As a result, managers find it increasingly difficult to react to their environment [MESC91]. The use of the organizational environment profile to assist in environmental scanning can help managers cope with dynamic business surroundings by gathering information on environmental influences expected to have a significant impact on the organization and by identifying emerging threats and opportunities resulting from environmental changes [MESC91]. Environmental scanning focuses on identifying indicators of potential environmental changes and thus is intended to alert the manager to potentially significant external influences before they have fully developed, thus allowing managers to anticipate problems before they occur rather than to react after they happen [NARA87]. While the inclusion of the organizational environment segment in the holistic profile enables information filtering systems to perform environmental scanning, this tool should be used in conjunction with other scanning technologies to insure the most thorough scanning possible.

Client-Type Profile

The client-type profile is essentially identical to the organizational environment profile, with the notable exception that it is developed from the client's perspective rather than from the organization's perspective. Many organizations

serve a particular type of client, and information that pertains to environmental factors that affect those clients are of vital concern to the organization. Unless the organization can keep abreast of factors that can benefit or harm their clients, the organization will be unable to advise and serve those clients effectively. The inclusion of the client-type profile, which has the same attributes as the organizational environment profile, helps to insure that the clients' welfare is accounted for.

Structure and Components of a Holistic Profile

Based on the findings detailed above, a definite structure suggests itself for each of the segments--personal, functional area, current project, organizational environment, and client type--of the holistic profile. This structure serves to guide the user in the selection of keywords for inclusion in the profile. The actual profiles that are generated should take into consideration all of the categories of keywords enumerated in each of the segments of the holistic profile, but even then a category that is significant to a particular user may be not be included. For this reason, each segment includes a set of user-defined categories that enable the user to tailor the holistic profile to his or her specific situation.

The personal profile appears in Table 4.1, the functional area profile appears in Table 4.2, the current project profile appears in Table 4.3, the organizational environment profile is shown in Table 4.4, and the client-type profile is shown in Table 4.5.

An Illustration of the Holistic Profile

An illustration of the holistic profile appears in the following scenario. David Robinson is a management accountant in a major telecommunications company. His personal profile, shown in Table 4.6, includes fiber optics, RISC machines, tennis, the Dallas Cowboys, the San Antonio Spurs, Texas Tech University, mergers of cable television and telecommunications companies, James Michener, his hometown Artesia, New Mexico, his current residence Phoenix, Arizona, and his favorite sports personality, David Robinson. His functional area profile, shown in Table 4.7, reflects

his position as a managerial accountant. His current project profile, shown in Table 4.8, reflects a project that involves compiling a survey of potential cable industry acquisitions. His organizational environment profile, shown in Table 4.9, is shaped by the factors pertinent to the telecommunications industry. The client-type profile does not pertain to this particular application, because the telecommunications industry does not serve a limited client base, and thus no client profile is provided.

It should be pointed out that Mr. Robinson's holistic profile was tailored to his needs either by utilizing the user-defined category or by omitting categories that are not applicable to his current situation.

Consider, for example, the list of possible newswire stories in Table 4.10. Although this list is obviously contrived, it serves to demonstrate how the filtering mechanism using a holistic profile is able to extract widely diverse stories that may be of interest to Mr. Robinson.

Obviously this list contains only those stories that were selected by the filtering process from the multitude of newswire stories. Many other stories did not match any keywords specified in the holistic profile and thus were not retained for Mr. Robinson's consideration. In addition, as was pointed out earlier, regardless of how carefully the filter is constructed it is inevitable that some non-significant information will pass through the filter. For example, while David Robinson is interested in mergers in general, he may have absolutely no interest in the merger between QVC and Paramount. Furthermore, while he is interested in how the NAFTA agreement affects the telecommunications industry, he may have little regard for Ross Perot's opinion on the topic of NAFTA. Thus, although those stories were retained for his consideration because of the inclusion of the keywords "merger" and "NAFTA" they are not of interest to him. However, the majority of the stories that were selected by the filtering process are of interest to the user and would not have been captured by existing user profiles. The holistic profile serves to broaden the scope of the stories that are presented for user consideration and thus makes an information filtering system a more useful tool.

Summary

By specifying a holistic profile made up of personal, functional area, current project, client type, and organizational environment keywords, this research has met the objective of broadening the scope of the profile to enhance its effectiveness. By increasing the purview of the profile, the filtering system can collect information of specific interest to the user, of interest to anyone serving in that particular capacity or on that particular project, and of interest to anyone in the overall organization. The inclusion of the organizational environment segment enhances the abilities of the filtering system, giving it the capability of acting as an environmental scanning device as well as a useful filtering device. These extensions to the profile make information filtering tools even more valuable in a corporate setting by providing decision makers with a more robust form of information technology to help them assess issues in an accurate, timely, and efficient manner.

Stage 2 – Specialized Templates

The purpose of this phase is to provide guidance to the user during the profile generation process. While the very structure of the holistic profile provides a degree of guidance by directing the user's keyword selection down appropriate paths, the provision of a template feature to suggest possible keywords to the user is an additional form of guidance intended to make profile generation as easy a task as possible. This requires the development of a knowledge base of sample profiles, or specialized templates, that serves as a listing of suggested keywords for the user in that particular context during the profile generation phase.

The specialized template feature consists of a knowledge base of possible templates for each of the segments of the holistic profile, with the exception of the personal profile segment, which is too individualized for the construction of a knowledge base of suggested keywords. However, for the functional area, current project, organizational environment, and client-type segments, potential templates will be provided. For example, if the user is employed in the telecommunications industry, during the construction of the organizational environment segment of his

or her holistic profile a sample template of potential keywords associated with the telecommunications industry will be provided by the system. The user can modify the template, adding or deleting keywords as best suits his or her particular situation. Templates designed for the organizational environment segment will be applicable to the client-type segment as well.

A sample template for the functional area of management accountant appears in Table 4.11. A sample template for the organizational environment or client type of telecommunications is shown in Table 4.12.

Stage 3 – Profile Expansion

The goal of this stage of the conceptual development is to provide a feature for profile expansion. Profile expansion is necessary because it is difficult for the user, even utilizing the provided templates, to select the exact set of words or word phrases that will result in the most successful filtering of information. This can be attributed to synonymy, which refers to the enormous variety of words that can be used to describe the same concept or object. Furnas, Landauer, Gomez and Dumais [FURN83, FURN87] discovered that across people, the same keyword is used only 10% to 20% of the time to describe an object. Thus, left to their own devices, people seldom select the keyword that results in the best payoff. One solution to this problem is to analyze the user's intermediate holistic profile and supplement the keyword selection with alternatives and synonyms. The user must then review the enhanced profile and remove irrelevant keywords, or the problem of polysemy is introduced. Polysemy refers to multiple meanings for a single word, which may result in extraneous data successfully passing through the filter.

There are several techniques that have been studied in the realm of profile expansion. Much research has been performed in the area of query expansion as it pertains to information retrieval. Other approaches have utilized the concept of thesaurus algorithms to provide as many synonyms as possible for each keyword, but this too may introduce some unwanted keywords. WordNet is another approach to keyword expansion that was investigated, as was latent semantic

indexing. The selected mechanism is a knowledge base of alternative keywords based on one of the above approaches.

An example of the keyword expansion process can be found in the above discussion. If the keyword "thesaurus" is one of the inputs to the profile expansion mechanism, additional keywords that might be suggested are "query expansion," "WordNet," "synonym generators," and "latent semantic indexing."

Stage 4 – Adaptive Profile

The goal of this stage of the conceptual development is to provide an adaptation mechanism for the profile based on usage patterns. It is unrealistic to expect users' interests to remain static over time. The information needs of any individual are continually evolving, and thus a mechanism must be provided to allow the profile to adapt to those changes. Frisse and Cousins [FRIS89] note that there is often an abrupt change in the desirability of information topics. Information that once seemed essential may suddenly become redundant, and the profile must be capable of adapting when such information becomes a nuisance [FRIS89]. Not only do some selected keywords become obsolete as user interests evolve, but some poorly selected keywords may be seldom used. When user interests change, the system must be able to notice that change, and further it must be capable of adapting in response to that change [SHET94]. Adapting to evolving user needs helps to make the system more useful and friendly [SHET94].

One method of adapting in response to user needs is "based on an analysis of user behavior patterns of which users are not aware" [FISC91, p. 68]. Based on these usage patterns the system can assist in modification of the profile to maintain its accuracy. This concept has its basis in Anderson's discourse on the Rational Analysis of Human Memory [ANDE90], which recognizes that past usage patterns help to predict future usage. In this study, a knowledge base will track the usage of individual keywords, and based on this usage will make recommendations to the user as to whether to retain or delete the keyword from the profile. The user has the option of heeding or ignoring the recommendations. This feature allows the

profile to adapt to the user's usage patterns and constantly evolve in line with the user's interests.

Summary

In summary, it provides clarification to review that basis for the conceptual development set forth in this chapter. This basis was derived not only by a review of existing research (Chapter II) but also through careful consideration of the purposes and goals of information filtering systems. The following can be asserted.

1. A comprehensive, or holistic, profile of a user's information interests is necessary before an information filtering system can perform to its fullest capabilities.
2. In order to provide a true management information systems tool this holistic profile must take into consideration the user's mental model, functional area, current project, organizational environment, and client type served.
3. Without some form of structure and guidance, the creation of the holistic profile will be a largely hit-or-miss affair; this structure and guidance can be provided in the form of pre-defined templates that assist in profile development.
4. Keyword-based information filtering systems are often less than totally successful because they fail to make provisions for synonymy; providing a synonym knowledge base helps to alleviate this problem.
5. A user's information needs evolve over time and so must the holistic profile; the inclusion of a passive learning mechanism allows the system to track keyword usage and alert the user when a keyword is no longer productive.

In concluding the discussion of the conceptual development, it is important to emphasize the distinguishing characteristics that set this research apart from existing research. There are a number of features that differentiate the concept of a holistic profile from conventional user profiles. The features offered by the holistic profile are unique in some cases because of their distinctiveness and in other cases because of their implementation. These features are detailed below.

1. An extensive literature search through technical journals of various disciplines, the Internet, and various research centers and universities revealed no conventional user profiles that cover the full spectrum of user information needs from a management perspective. The majority of existing work in information filtering and user modeling focuses on user interests and excludes functional area, current project, organizational environment, and client type information needs. As such, their application as a management information systems tool is limited.
2. Although some notions have been adapted from existing work in information filtering systems, this work is unique in that it offers the entire set of features. For example, while some systems provide limited features to assist the user in profile development, others include provisions to counteract synonymy, and still others include a profile that adapts to changing user interests, no systems have been found that include the combination of a comprehensive management-based profile, a feature that provides structure and guidance for profile development, a feature to deal with synonymy, and an adaptive profile.

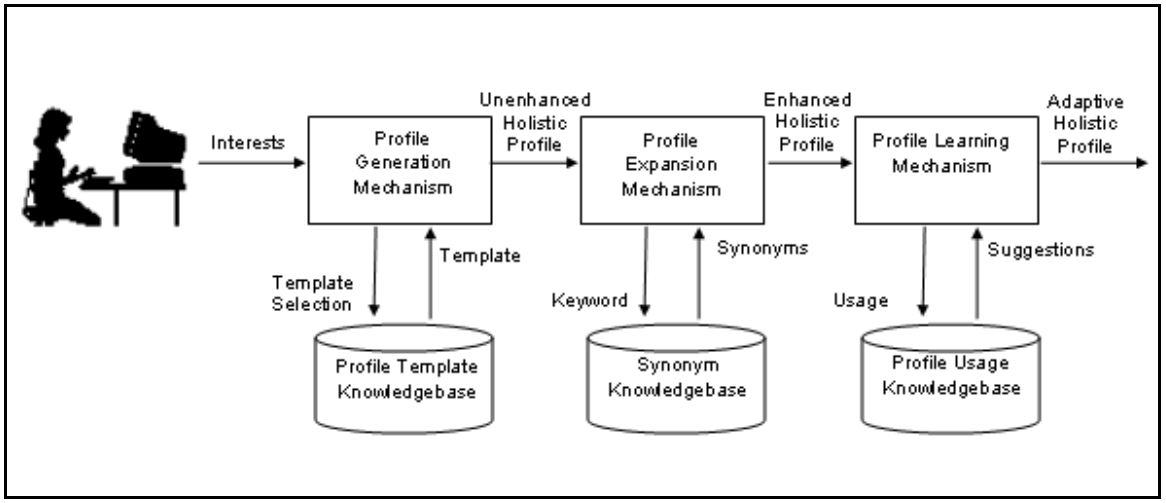

Figure 4.1. Holistic Profile System Architecture.

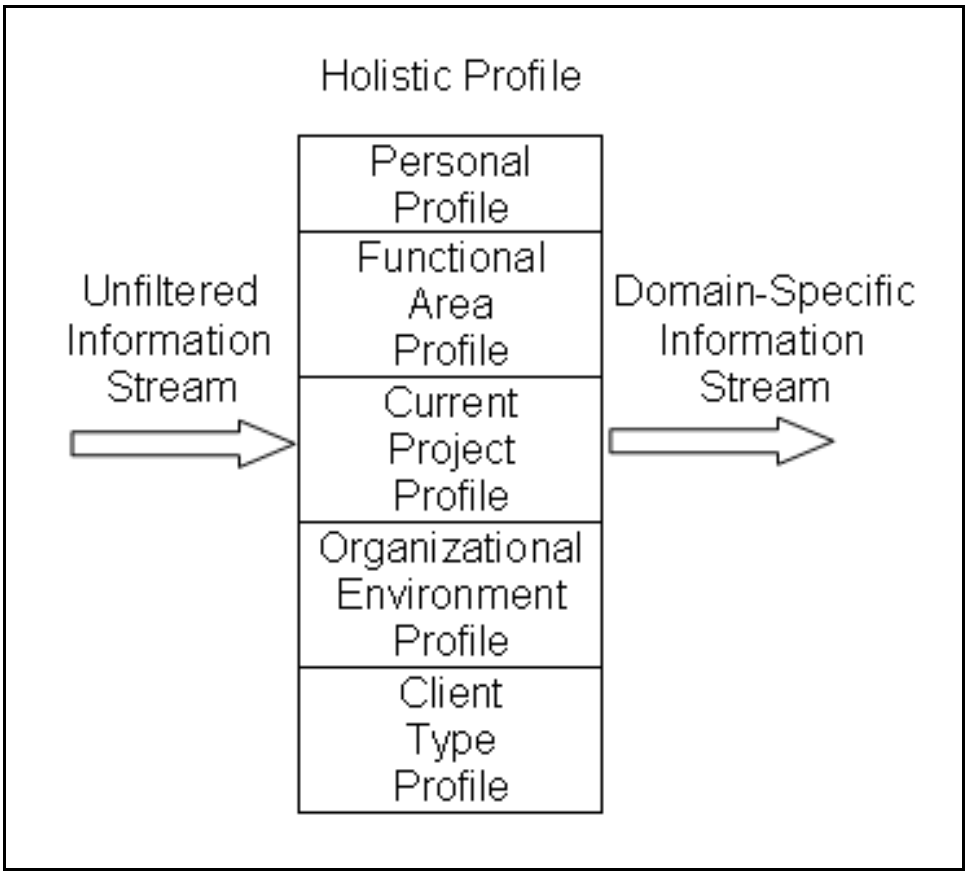

Figure 4.2. Filtering with a Holistic Profile.

POSITION: AUDITOR, INTERNAL DEPARTMENT: Accounting

SCOPE OF JOB:
Perform professional internal auditing activities in the recording an reporting of financial transactions and budgetary control for assigned company departments; review accounts payable payroll and other fiscal records; prepare financial report statements and analyses.

This is a management-level classification in the accounting function. Individual must perform the full range of professional accounting and internal audit activities in the company. He or she must have a thorough knowledge of accounting systems and corporate procedures, and the ability to exercise independent judgment within established systems and programs.

SUPERVISION STATUS:
- Receive supervision from the Director of Finance. Functional or technical supervision provided by the Senior Accountant/Auditor.
- Provide supervision to technical and clerical personnel.

RESPONSIBILITIES: (May include, but are not limited to):
- Assist various departmental personnel with budget analysis, expenditure comparisons, document preparation and other accounting-related research and analysis activities.
- Assist in annual budget preparation by function and for entire company.
- Audit business activities to insure compliance with governmental ordinances, regulations and legal prescriptions.
- Prepare audit papers from financial statements and various reports; review federal, state, and local forms for internal accounting and auditing purposes.
- Monitor and participate in the reconciliation of the general ledger and subsidiary accounts.
- Review the status of various audit programs and reports.
- Recommend and implement changes in accounting/audit systems and procedures.
- Monitor budget activities for various company departments and projects to determine if funds are available and expenditures properly classified; research and analyze transactions to resolve budget problems; provide analysis of available funds at management's request.
- Prepare and review account balance projections with other departments.
- Assist with the preparation of the company's annual financial report.

MINIMUM REQUIREMENTS:
- Apply accounting and audit principles and practices.
- Recognize principles of internal/external financial administration, including budgeting, reporting, taxation, and revenue.
- Have working knowledge of modern office practices, procedures, methods, and equipment.
- Follow updated supervision, training and performance evaluation techniques.
- Operate modern, automatic and mechanical accounting equipment and personal computers.

STANDARD DUTIES:
- Examine and verify financial documents and reports.
- Communicate effectively, orally and in writing.
- Prepare a variety of financial/audit statements, reports and analyses.
- Establish and maintain effective working relationships with various departments.
- Supervise, train, and evaluate clerical and technical personnel for the Auditing Department.

BACKGROUND:
College degree with emphasis in business administration and statistics; four years' audit experience with emphasis in accounting.

Figure 4.3. Sample Job Description [WOLF88, p. 115].

Table 4.1. Personal Profile.

Keyword Category	Definition	Example Keywords
Technical Interests	Technical areas of interest to the user.	Intel Pentium Microprocessor.
Hobbies	Hobbies for which current information is desired.	HO Railroads
Sports teams	Sports teams that the user follows.	Texas Longhorns
Alma mater	The school that the user graduated from.	University of Texas
Current events	Current news stories that affect the user.	Airline strike
Authors	Authors that the user enjoys.	Raymond Feist
Hometown	Details that pertain to the user's roots.	Saratoga Springs, New York
Current residence	Details that pertain to the user's residence.	Dallas, Texas
Personalities	Personalities in whom the user is interested.	Lee Iacocca
User-defined categories	Miscellaneous interests that do not fit neatly into other categories.	Airfare reductions

Table 4.2. Functional Area Profile.

Keyword Category	Definition	Example Keywords
Technical skills	Technical skills needed to perform duties of functional area.	Computer literacy, accounting software, word processing
Managerial skills	Skills needed to manage subordinates	Performance appraisal
Interpersonal skills	Skills needed to relate to and communicate with superiors, peers, and subordinates.	Communication skills, leadership skills, group problem-solving techniques
Professional development	Skills needed to update the user's professional skills and knowledge.	Computer skills, accounting practices
Professional associations	Professional or industry associations with which a professional should be involved.	American Institute of Certified Public Accountants, Association for Computing Machinery
Regulatory issues	Regulations issued by governmental or licensing bodies that directly affect a profession.	Taxation regulations, certification requirements
Career progression	Skills that must be developed before a promotion is considered.	Managerial ability
Evaluation shortcomings	Areas needing improvement from performance appraisals.	Writing and speaking skills
User-defined categories	Miscellaneous interests that do not fit neatly into other categories.	

Table 4.3. Current Project Profile.

Keyword Category	Definition	Example Keywords
Project Goals or Objectives	Overall focus and goals of the project. Includes such factors as scope, time and cost parameters, priorities, environmental, technological, and operational constraints, milestones, and control considerations.	Upgrade corporate computing system
Project Basis or Background	Premise or background information on which the project is based.	486DX2, Pentium, PowerPC
Technical Requirements	The technical requirements associated with the project.	RAM, Fixed disk capacity
Resource Requirements	The number and skill level of personnel involved in the project or the project results; the cost of materials required as well as their availability.	Knowledge of Windows 3.1; Availability of PowerPC chip
Stakeholder Concerns	Includes concerns of stakeholders such as clients, users, managers, financiers, suppliers, and management.	Concerns of venture capitalists
Market Considerations	Progress of competition, product differentiation, economic environment.	Research into additional uses for digital signal processor chip
Project Management Priorities	Includes scheduling, cost control, quality control, scope management, contract management, resource management, and interface management.	Cost of materials is increasing; mean time between failures of vital components is unpredictable
Financial Requirements	Factors that influence project financing or costs.	Price cuts in the PC market
Return-on-investment Factors	Factors that could influence the projected return-on-investment.	Decrease in the price of Bell-Atlantic stock
User-defined categories	Miscellaneous interests that do not fit neatly into other categories.	

Table 4.4. Organizational Environment Profile.

Keyword Category	Definition	Example Keywords
Customer Base and Marketplace	The group or type of customer that the organization serves.	Executive computer users
Industry	The industry type and major products that the organization produces.	Telecommunications products
Socioeconomic and Demographic	Socioeconomic factors that affect demand for the organization's products.	Population explosion, unemployment, increasing affluence
Competition and Corporate Environment	Industries that produce the same or substitute products.	AT&T
Manpower and resources	Availability of qualified personnel and organizational resources	Shortage of technically competent employees
Technological	Technological factors that affect the organization's performance and success.	New manufacturing methods, e.g silicon wafer etching
Geopolitical	Geopolitical factors that affect the organization's markets and expansion.	Asian-Pacific Economic Cooperation forum, EEC, NAFTA
Governmental	Regulations and trade barriers that affect an organization's markets such as import/export laws, taxation issues, business law, patent and trademark law.	Telecommunications deregulation, regulation of multinational corporations
User-defined categories	Miscellaneous interests that do not fit neatly into other categories.	

Table 4.5. Client Type Profile.

Keyword Category	Definition	Example Keywords
Customer Base and Marketplace	The group or type of customer that the organization serves.	Executive computer users
Industry	The industry type and major products that the organization produces.	Telecommunications products
Socioeconomic and Demographic	Socioeconomic factors that affect demand for the organization's products.	Population explosion, unemployment, increasing affluence
Competition and Corporate Environment	Industries that produce the same or substitute products.	AT&T
Manpower and resources	Availability of qualified personnel and organizational resources	Shortage of technically competent employees
Technological	Technological factors that affect the organization's performance and success.	New manufacturing methods, e.g. silicon wafer etching
Geopolitical	Geopolitical factors that affect the organization's markets and expansion.	Asian-Pacific Economic Cooperation forum, EEC, NAFTA
Governmental	Regulations and trade barriers that affect an organization's markets such as import/export laws, taxation issues, business law, patent and trademark law.	Telecommunications deregulation, regulation of multinational corporations
User-defined categories	Miscellaneous interests that do not fit neatly into other categories.	

Table 4.6. Example of a Personal Profile.

Keyword Category	Keywords
Technical Interests	Fiber optics, RISC
Hobbies	Tennis
Sports teams	Dallas Cowboys, San Antonio Spurs
Alma mater	Texas Tech University
Current events	Corporate mergers, telecommunications, cable television
Authors	James Michener
Hometown	Artesia, New Mexico
Current residence	Phoenix, Arizona
Personalities	David Robinson
User-defined	Airline strikes

Table 4.7. Example of a Functional Area Profile.

Keyword Category	Keywords
Technical skills	Financial statement analysis, budgeting/profit planning
Managerial skills	Performance appraisal
Interpersonal skills	Communication skills, leadership skills, group problem-solving techniques
Professional development	Computer-assisted accounting
Professional associations	American Institute of Certified Public Accountants
Regulatory issues	Tax compliance, Corporate taxation, Certified Management Accountant requirements
Career progression	Management skills
Evaluation shortcomings	Attain CMA certificate

Table 4.8. Example of a Current Project Profile.

Keyword Category	Keywords
Project Goals or Objectives	Introduce fiber optic network to rural subscribers
Project Basis or Background	Need to extend availability of fiber optics to all client base
Technical Requirements	Knowledge of fiber optic circuit equipment
Resource Requirements	Sales ability, experience with fiber optics, pricing of fiber optic cable
Stakeholder Concerns	Rural customer base
Market Considerations	Southern Bell use of fiber optics for home wiring
Project Management Priorities	Costs of transmission of information
Financial Requirements	Costs of replacement of copper wire
Return-on-investment Factors	Standardization of connecting devices

Table 4.9. Example of an Organizational Environment Profile.

Keyword Category	Keywords
Industry	Telecommunications
Socioeconomic and Demographic	Distributed families, collaborative workgroups, image-intensive communications
Competition and Corporate Environment	Mergers, acquisitions, and divestments of competitors, developments in communications media
Manpower and resources	Shortage of fiber optic cable suppliers
Technological	Breakthroughs in telecommunications and cable television, fiber optics, digital telecommunications
Geopolitical	NAFTA, Telefonos de Mexico, TELMEX
Governmental	FCC regulations, FTC regulations, corporate mergers, regulation of multinational corporations

Table 4.10. Newswire stories and matching profile segment.

Topic of Newswire Story	Profile/Category
American Airlines strike	Personal/User-defined
Bell Atlantic and TCI merger	Personal/Current events Organizational/Industry Organizational/Competition and Corporate Environment Organizational/Governmental
Rash of sports-related attacks such as that on tennis star Monica Seles	Personal/Hobbies
Dallas Cowboys' chances of a repeat appearance in the Super Bowl	Personal/Sports teams
California earthquake with tremors felt as far away as Phoenix	Personal/Current residence
Comparison between basketball centers Hakeem Olajuwon and David Robinson	Personal/Personalities
Analysis of the demographics of widely-separated families in the 1990's	Organizational/Socioeconomic and Demographic
Editorial on the merger between QVC and Paramount	Personal/Current events Organizational/Competition and Corporate Environment Organizational/Governmental
Reaction of the Clinton administration to proposed mergers between telecommunication and cable industry giants	Personal/Current events Organizational/Industry Organizational/Competition and Corporate Environment Organizational/Governmental
PowerPC: an alternative to Intel	Project Management/Basis/Background
Discussion on the importance of properly conducting performance appraisals in the workplace	Functional Area/Managerial skills
Impact of the 1993 Revenue Reconciliation Act on corporate taxation	Functional Area/Regulatory issues
Ross Perot's opposition to the approval of NAFTA	Organizational/Geopolitical

Table 4.11. Functional Area Profile Template for a Management Accountant.

Keyword Category	Template Items
Technical skills	Financial statement analysis, budgeting/profit planning
Managerial skills	Performance appraisal
Interpersonal skills	Communication skills, leadership skills, group problem-solving techniques
Professional development	Computer-assisted accounting
Professional associations	American Institute of Certified Public Accountants
Regulatory issues	Tax compliance, Corporate taxation, Certified Management Accountant requirements
Career progression	Management skills
Evaluation shortcomings	Attain CMA certificate

Table 4.12. Organizational Environment Profile Template for Telecommunications.

Keyword Category	Template Items
Industry	Telecommunications
Socioeconomic and Demographic	Distributed families, collaborative workgroups, image-intensive communications
Competition and Corporate Environment	Mergers, acquisitions, and divestments of competitors, developments in communications media
Manpower and resources	Shortage of fiber optic cable suppliers
Technological	Breakthroughs in telecommunications and cable television, fiber optics, digital telecommunications
Geopolitical	NAFTA, Telefonos de Mexico, TELMEX
Governmental	FCC regulations, FTC regulations, corporate mergers, regulation of multinational corporations

CHAPTER V

HOLISTIC PROFILE FILTERING SYSTEM PROTOTYPE DESIGN

Introduction

The knowledge-level and symbol-level principles set forth earlier, as well as other decisions made during conceptual development, led to a preliminary design for the holistic profile prototype system. This chapter describes the design of each of the components of the prototype, beginning with the generic filtering system, continuing with the conventional profile, and concluding with the various modules that comprise the holistic profile filtering system. The prototype design issues surrounding the holistic profile filtering system are based on the symbol-level principles discussed in the previous chapter. Because the information filtering system and the conventional profile are included for comparison testing purposes only, the symbol-level concepts upon which they are based are not discussed in this document.

Information Filtering System

Because none of the existing information filtering systems provided the range of capabilities required by the comparison study, it was necessary to develop a generic filtering system. An information filtering system filters incoming streams of data through a user profile, discarding data that fail to match the profile and retaining only that information relevant to a user's expressed interests. The overall design of the prototype filtering system is shown in Figure 5.1.

Requirements of the Information Filtering System

There are several issues involved in the design of the information filtering system. First, the system must be capable of accepting text-based documents and performing keyword matching on those documents. Second, it must be designed so that the filtering system is independent of the profile, which will allow the system to utilize either the holistic profile or the conventional profile as the source of the

keywords. In Figure 5.1 the symbol labeled "user profile" represents either the conventional user profile or the holistic profile.

Information Filtering System Design Decisions

In addition to the general requirements discussed above, more specific requirements must be dealt with. The other system requirements include provisions for:

- searching for keywords;
- displaying retained documents;
- tracking keyword performance; and
- tracking recall variables.

Searching for Keywords

The first requirement that must be resolved deals with the method by which the filtering mechanism searches an article for the keywords. The system should permit the use of wildcards in the search mechanism, because it is not always enough to search for an exact match on a string. The search mechanism must also have the ability to handle hyphenated words, words containing an apostrophe, or words followed by any of the various punctuation symbols. Further, the system must allow the user to specify a key phrase rather than just a keyword, a design requirement that presents additional problems when the document being searched contains a key phrase that begins on one line and is continued on the following line. All of these issues present different implementation problems, but their resolution is necessary to provide a fully functional filtering system.

Other design issues revolve around recording which keywords are responsible for the retention of a document. Although a match on a single keyword is enough to determine that an article should be retained by the filter, it is not enough to find that first match and then stop searching. All of the keywords in the profile must be examined to determine if they appear in the document and play a role in its final disposition. Thus, the system must search for and record each keyword that appears in a retained document. In addition, each line in a document

that contains a keyword must be flagged. That flag makes it possible for the display mechanism to determine which lines contain keywords.

Displaying Retained Documents

In order to utilize any filtering system, the user must be able to read and assess the documents provided by it. In the case of the holistic profile prototype, one of the design issues pertains to ease of use issues. In order to make the system more user friendly, and to provide the most advanced viewing mechanism possible, the document display mechanism should be capable of highlighting in yellow every line containing a keyword. This achieves the same effect as using a fluorescent highlighter to accent each interesting line in a paper-based document. The user's attention will be drawn to those lines that contain keywords, allowing him or her to more quickly assess the usefulness of the article.

Tracking Keyword Performance

After reading each document, the user must also be able to evaluate or rate the article's content. This step is necessary because this datum is required in order to track the performance of each of the profiles. One of the design issues revolves around the evaluation screen that is displayed immediately after the user concludes scanning the current document. The evaluation screen requires the user to assess the document with regard to how well it corresponds to the user's interests. The evaluation screen should be designed such that the user can review the article if necessary before evaluating it. This calls for a mechanism to allow the user to temporarily leave the evaluation screen and redisplay the document in question. Further, the user also should be given the opportunity to archive a document for future display on a document-by-document basis rather than arbitrarily saving or deleting each article.

Tracking Recall Variables

A final issue that became apparent during the design of the information filtering prototype is the need for the user to review not only the documents that are

retained by the information filter, but also those documents that are bypassed. One of the measures by which the effectiveness of an information filtering system can be determined is recall, which was defined earlier as the percentage of available relevant documents that are actually retained. The user evaluations of the retained documents indicates which of those are relevant. However, in order to determine the total number of relevant documents that are available, the user must also evaluate those documents that are bypassed by the system. The number of documents that are bypassed but still rated as interesting added to the number of documents that are retained and rated as interesting provides the number of available relevant documents, allowing recall to be calculated.

The Conventional Profile

The profile is a list of keywords that describes the user's information interests. It is the component of the information filtering system that determines whether a document is retained or discarded. If the document contains one or more of the keywords in the profile, then it is retained for the user's consideration.

Requirements of the Conventional Profile

The detailed design for the conventional profile is less complex than that of the filtering system. The major requirement for the conventional profile is that it fits seamlessly into the filtering system, and that it provides the filtering system with a list of keywords describing user interests. The conventional profile must be representative of existing systems, and therefore should provide very little user guidance such as extensive keyword categories or templates. The user simply enters a list of keywords via an input screen.

Conventional Profile Design Decisions

Not only must the general requirements of a conventional profile be provided for, but more specific requirements remain. These specific requirements include:

- providing input field labels;

- personalizing user-defined categories;
- allowing multiple keywords;
- providing control buttons; and
- maintaining the conventional profile.

Providing Input Field Labels

The design of input screens for the conventional profile will provide a standard on which all subsequent input screens will be based. The standard input screen for profile generation consists of labeled input fields and a set of form controls. The input fields are labeled with the keyword category. In the case of the conventional profile, the keyword categories consist of vague and/or sweeping categories such as "people," "companies," "product names," "action verbs," etc. The choice of categories is modelled after the profile guidelines provided by the NewsEDGE[1] filtering system [NEWS93].

Personalizing User-Defined Categories

The standard input form will include three or more user-defined categories to accommodate keywords that do not fit neatly into other categories. Each user-defined category will be designed so that the user can replace the existing label with a new label. This feature makes it necessary to save not only the keywords, but also the labels.

Allowing Multiple Keywords

When entering keywords for a particular category, the user should have the option of entering multiple keywords separated by commas. As each field is completed, the entries should be stored in a temporary file, and only saved to the conventional profile when the user requests it. This will allow the user to modify, add, or delete keywords from fields before they are saved. Duplicate entries are unnecessary and therefore will be removed from the profile.

---

[1]NewsEDGE is a registered trademark of Desktop Data, Inc.

Providing Control Buttons

Each input form should also include the standard control buttons to save all entries, clear all entries, invoke the help utility, go to the next screen, and quit. In addition to these standard buttons, forms must also include a button to undo the clear operation, i.e., to restore the screen to its former state in case the user clears the form by mistake.

Maintaining the Conventional Profile

Finally, the user must be able to modify, add, and delete entries in the conventional profile. In conjunction with this requirement is the ability to print the conventional profile so that the user has a paper copy to refer to.

Many of the design decisions arrived at while designing the conventional profile, such as decisions made regarding screen design, also apply to the holistic profile.

The Holistic Profile

In this section the term "holistic profile" is used to refer to the holistic counterpart of the conventional profile. The design of the holistic profile filtering system is shown in Figure 5.2. With regard to functionality, there are few differences between the holistic profile and the conventional profile. However, as was discussed in the previous chapter, the holistic profile greatly expands the scope of the user profile.

Requirements of the Holistic Profile

The primary difference between the holistic profile and the conventional profile is the number of segments that make up the holistic profile. While the conventional profile is made up of a single segment of general user interests, the holistic profile is highly structured and segmented into specialized groupings of user interests. This helps to satisfy both the first and the fourth symbol-level principle, which call for a segmented architecture in which all pertinent areas of user interests are identified and accounted for.

Holistic Profile Design Decisions

More specific requirements for the holistic profile include:

- providing multiple input forms;
- selecting template items; and
- maintaining the holistic profile.

Providing Multiple Input Forms

Because the holistic profile is made up of multiple segments, a single generic input form is no longer enough. Each segment requires an individual input screen, all of which are designed in a fashion similar to the conventional profile input screen. Keyword labels will differ from screen to screen, as will the templates. Because there are multiple input screens, an additional button must be added to complement the "next" button. The additional button will allow the user to return to the previous input screen if desired.

Selecting Template Items

One of the enhancements of the input screens for the holistic profile is necessitated by the fact that the holistic profile provides templates to assist the user in profile development. While no templates are provided for the personal profile segment of the holistic profile, the functional area segment, the ongoing project segment, the organizational environment segment, and the client type segment all feature the template provision. While the template feature will be discussed in the next section, and implementation issues are the subject of the following chapter, templates cannot be provided without enhancements to the standard input screen. At this point it is sufficient to say that each input field must be designed to accept not only typed entries, but also entries selected from the templates. The standard methods for editing must remain available, despite the special nature of the field.

Maintaining the Holistic Profile

A final design issue revolves around profile maintenance. Like the conventional profile, the holistic profile must allow the user to make modifications, additions, and deletions to the holistic profile. Again, the system must also allow the user to print a copy of the profile for review.

Template Module

The template feature is designed to provide assistance to the user in the form of keyword suggestions. For example, if the current input field is "technical skills" on the functional area input form, an accountant will be provided with a specialized template of technical skills suitable for an accountant.

Requirements of the Template Module

There are two primary requirements that must be satisfied by the template feature. First, the templates must be domain specific. That is, if the user indicates that he or she works in a particular industry, then when entering user interests in the organizational environment segment of the profile, a template for that industry should be available. This design issue can be traced to the third symbol-level principle, which calls for domain-specific templates to provide user guidance. Second, the user must be allowed to decide how heavily he or she wants to rely on the supplied templates.

Template Module Design Decisions

The primary design decisions relate to the following:

- determining the correct template; and
- interacting with the template feature.

Determining the Correct Template

As noted in the preceding section, many of the input forms for the holistic profile provide domain-specific templates to assist the user in profile development. The requirement that the templates be "domain specific" presents some difficulty because the domain varies from input form to input form. For instance, the domain

for the input form might be "accountant" while the domain for client type is "construction." In order to facilitate the determination of domains, the holistic profile input process is initiated by an overview form that requires the user to specify the job title, the project type, the industry type, and the client type. The user response is used to establish the domain for each of the forms, and is used to select not only the appropriate template, but also the appropriate synonym set.

Interacting with the Template Feature

Care must be taken in the design of the template feature. Suggestions must not be forced on the user; e.g., a bad design would fill the input field with suggestions and require the user to delete those that are not to his or her liking. The suggestions must be more unobtrusive, but easily available for the user to call upon. The system must provide the user with a list of multiple suggestions for a keyword category, and allow the user to select among those suggestions, choosing one or more of the selections. The user must also be able to enter keywords of his or her own choice in addition to or instead of the template offerings.

As with the topics covered previously in this chapter, additional issues surrounding the template feature will be discussed in the implementation chapter.

Synonym Module

The synonym module is intended to assist the user in keyword selection by providing supplemental or alternative keywords to those entered by the user during profile input.

Requirements of the Synonym Module

Like the templates, the synonyms must be domain specific as specified in the fifth symbol-level principle. Because they are domain specific, the correct synonym set can be determined by user responses to the overview screen, as discussed in the previous section. If the user responds with "accountant" for the job title field on the overview screen, when the user is entering keywords in the functional area input

screen the profile expansion mechanism will provide a synonym set designed specifically for an accountant.

Synonym Module Design Decisions

Additional requirements for the synonym module include:

- making synonyms available; and
- selecting synonyms.

Making Synonyms Available

Synonyms are not provided to the user immediately upon data entry. As noted earlier, user entries are not permanently saved when the user exits a field. Instead, as each field is completed the entries are added to a temporary file, and only saved to the conventional profile when the user requests it. When the user directs the system to save the keywords on the input form, the system will examine each entry to determine if synonyms are available. If so, the user will be alerted that alternatives exist and the system will display the keyword entered along with a list of synonyms that are available for that keyword.

Selecting Synonyms

The user must be given the option of selecting zero or more of the synonyms, and must also be allowed to choose whether the synonyms will be used in place of, or in addition to, the original entry. The synonym list can be used not only to assist the user in keyword selection, but also to expand common acronyms such as CPA, or certified public accountant.

Passive Learning Module

The passive learning mechanism must be designed to track keyword performance with respect to how often each keyword contributes to the retention of a document, as well as the type of evaluations that are associated with articles retained by the keyword. By so doing the system takes on a pseudo-adaptive quality that makes it possible to suggest profile modifications, as called for in the second symbol-level principle.

Requirements of the Passive Learning Module

File design is an important issue that must be resolved at this point. The type of information necessary to track keyword performance must be stored with the keyword. It is also essential to determine the most effective form of adaptation.

Passive Learning Module Design Decisions

As noted above, the two primary design issues include:

- tracking keyword performance; and
- selecting the adaptation technique.

Tracking Keyword Performance

In order to track keyword performance several key variables must be associated with each keyword. Not only must each type of profile include the keyword category and keyword, but each record must also include the number of documents retained due to the keyword, the total score achieved by the keyword based on user ratings of the retained documents, and the number of documents retained by the keyword that have been read and evaluated. By tracking this information the passive learning mechanism can evaluate the performance of each keyword.

Selecting the Adaptation Technique

During the design phase it was again considered whether to make the learning mechanism an automatic feature or a passive feature. An automatic feature would continuously evaluate keyword performance, weeding out those that have a sub-par performance. Acting as an intelligent agent, the mechanism would arbitrarily decide the fate of each keyword. The alternative is a passive mechanism, which tracks the performance and, when invoked by the user, alerts the user to any keywords that are not performing in a satisfactory manner. In this way the user makes the final decision about the disposition of a keyword. The latter approach was selected based on the argument that the user, rather than an arbitrary

algorithm, should make the final determination about whether to retain or delete keywords from the holistic profile.

Summary

The design decisions reached during the detailed design phase are based on the symbol-level principles and provide direction for the implementation phase. While this chapter described the various considerations that contributed to the overall design of the system, the following chapter will describe how those design decisions were ultimately implemented.

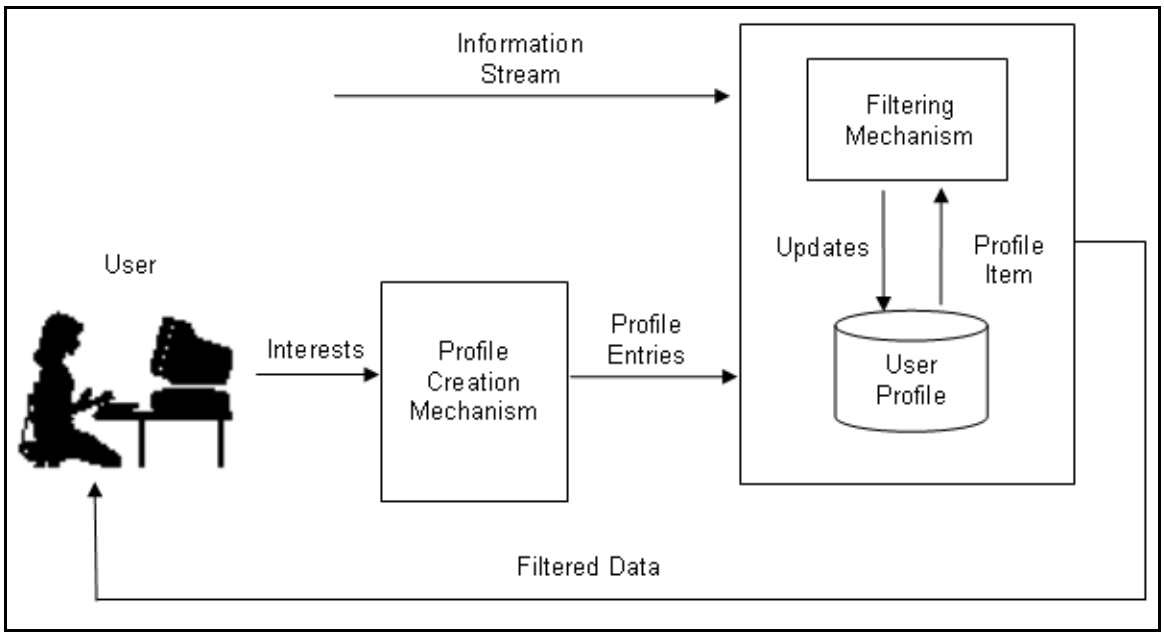

Figure 5.1. Design of an Information Filtering System.

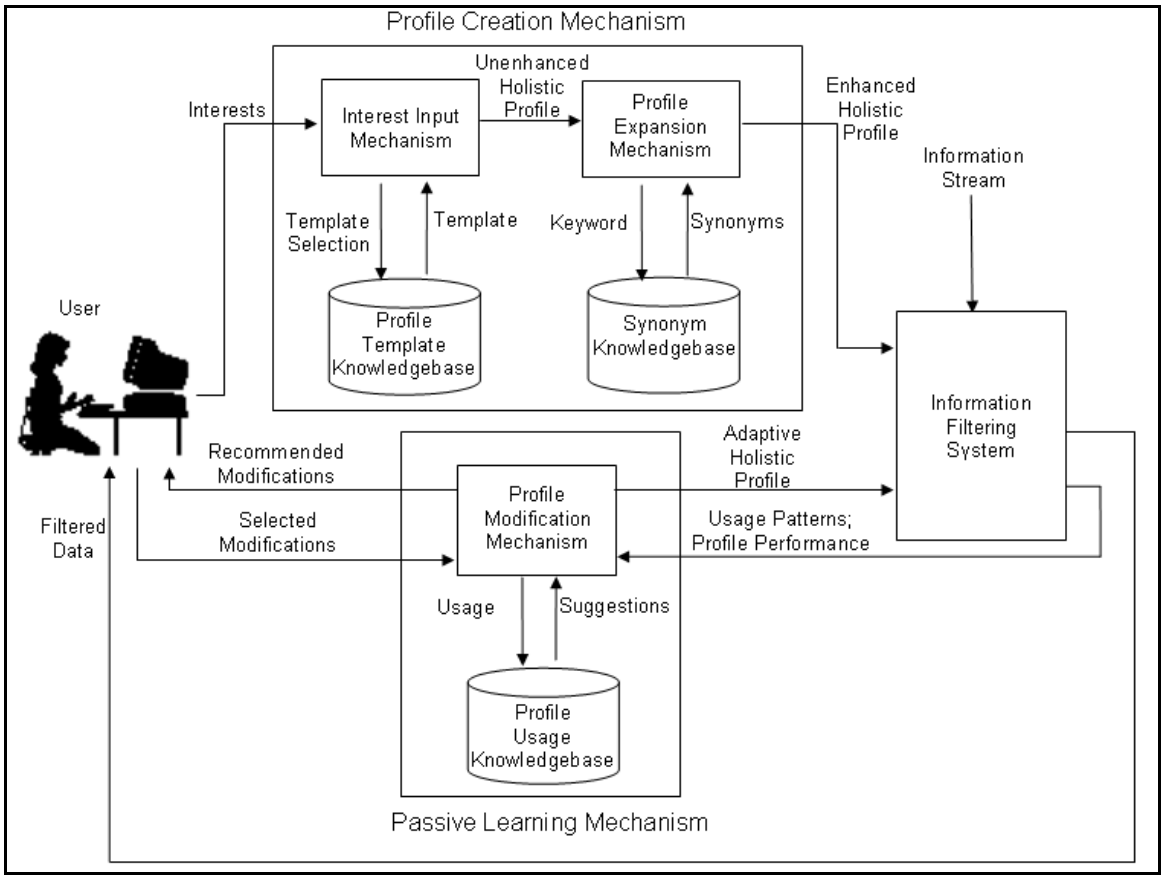

Figure 5.2. Design of the Holistic Profile Filtering System.

CHAPTER VI

IMPLEMENTATION

Introduction

The implementation of the prototype is an essential step in demonstrating the soundness of the conceptual design. Without a working model, any claims made about the new design remain unsubstantiated. Before the implementation can be undertaken, however, a suitable tool must be identified and evaluated.

Tool Selection

The first step in tool selection is the determination of the most suitable class of tools. The tools that are best suited for the implementation of an information filtering system include standard programming languages and relational database packages. The decision was made to use a database package because of its ability to maintain keyword tables, its assistance in the design of input forms, its ability to import text files, and its ability to search those text files for keywords or key phrases. While many software packages offer tools to assist in applications development, the built-in features of most database packages provide greater convenience.

After selecting the type of tool for prototype implementation, it was necessary to decide on a specific tool. After discussions with experienced developers regarding appropriate database packages, it became apparent that the consensus was Microsoft Access.[2] This database package offers all of the capabilities noted in the preceding paragraph. Access also provides an event-driven approach to application design, and its own dialect of Object Basic, a programming language also referred to as Visual Basic for Applications [JENN94].

Given the prototype design described in the previous chapter and the capabilities and limitations of Access, the prototype development entered the

---

[2]Microsoft Access is a trademark and Microsoft is a registered trademark of Microsoft Corporation.

implementation phase. The implementation began with the information filtering system, continued with the conventional profile, then moved on to the holistic profile and its various features including the template module, the synonym module, and the passive learning module.

Information Filtering System

The information filtering system is intended to demonstrate the performance difference between the conventional user profile and the holistic profile. It must accept an incoming stream of data, compare it to a user profile, and retain only those data that match the user's expressed interests. It must also be capable of displaying the retained data for the user.

Filtering information presents several special problems that are unique to this portion of the prototype. The primary concern is to automate as many of the tasks as possible. Each operation must also be as robust as possible. While many operations, such as importing files, can be readily performed manually in Access, it is an entirely different matter to implement the process in Access Basic.

The filtering system must be capable of:

- copying text files from the daily distribution diskette;
- importing text files into Access database tables; and
- filtering text through the profile.

Copying Text Files

Each participant was provided with a diskette containing a set of current news items on a daily basis. The task of copying text files from the diskette to the user's Access subdirectory was originally intended to be a manual task that would be performed by the researcher when the diskette was delivered daily. However, it is not reasonable to expect the user to always be able to relinquish his or her computer to the researcher immediately upon his arrival. Therefore it was necessary to automate the task of copying news articles. Access Basic provides a command called Shell, which will temporarily leave Access and run any executable program with an .EXE, .COM, .BAT, or .PIF extension. By providing each distribution

diskette with an executable file called LOAD.BAT to copy the files from the diskette drive to the Access subdirectory, it was possible to use the Shell command to invoke the load routine and copy the files automatically. The load routine is initiated when the user clicks on the import button. Figure 6.1 shows the menu from which the import action is initiated.

Importing Text Files

After giving the user the option of loading files from diskette, the import button invokes the import routine. This routine is responsible for importing each news article from a text file and converting it into a format that can be stored in an Access table. The import routine first creates a new table in which to import the text data. This is done by using the Access Basic CopyObject command to make a copy of a table template, which consists of a table structure but no data. The routine next uses the Access Basic TransferText routine to import the file. The TransferText routine requires a specification name, the table name into which data are being imported, and the file name from which data are being exported. The specification name is the name of the import/export specification that informs the TransferText command of the structure or format of the incoming data. For example, in importing data from text files the import/export specification indicates that the data are in fixed width fields, that they are of type text, and that the fields can contain up to 125 characters. The command converts each line of the text file into a record in the specified database table. This operation is performed for every text file that is provided for the current day, converting each news article into a database table.

Filtering Text

The implementation of the actual filtering process is also a challenging undertaking. The primary issue is the technique by which each article is searched for keywords. Access Basic provides the FindFirst and FindNext commands to find the first or next record that satisfies a specific condition, in this case the inclusion of a particular search string. The Access Find commands presented some difficulties

because of their restriction to dynasets, the structure of the search string, and the inability to locate key phrases.

Restriction to Dynasets

One peculiarity of Access Basic is that the FindFirst and FindNext commands cannot be used to search a table, but can only be used with dynaset-type recordsets. A dynaset is a dynamic set of records that results from a query on an underlying table. While this presents no major problems, it necessitates that any tables that are to be searched are opened as dynasets rather than as tables.

Structure of Search Strings

Other oddities associated with the search process include the fact that in order to find a keyword in a line of data, the search string must be preceded and followed by a wildcard character, or asterisk. This makes it possible for the search to match a line of text that contains any character string, represented by the asterisk, followed by the search string, followed by any character string, again represented by the asterisk. For example, to locate the keyword "construction," the search string should be "*construction*." If the wildcards are omitted and the search string is "construction" then it will only match lines containing the single word "construction." The system must also be able to locate the keyword only, rather than the keyword embedded in another word. This can be accomplished by following the first asterisk with a blank. For example, the search string "* construction*" will match "construction" but not "reconstruction." In order to locate a search string even when it is part of a larger string, the user can append a wildcard. This allows the system to use a search string such as "* construct*" to locate strings like "constructed," "construction," and "constructive." However, if the user wishes to locate whole words only, then the wildcard is omitted and the filtering mechanism establishes a search string such as "* construct[-?]*." This will locate the search string followed by a blank, an exclamation point, a quotation mark, a percent sign, an ampersand, an apostrophe, a parenthesis, a comma, a period, a hyphen, a slash, a colon, a semicolon, or a question mark--in short, any member of

the ASCII character set that is included in the range from a blank space (decimal code 32) to a question mark (decimal code 63).

Inability to Locate Key Phrases

The final problem associated with searching stems from the use of key phrases rather than keywords. If the phrase is contained in a single line in the table then Access is able to locate it. However, if the phrase begins on one line and is continued on another, then Access is unable to find a match, and incorrectly reports that the key phrase is not contained within the table. In order to locate such strings, the filtering system is implemented in such a way that it searches for the initial word in the key phrase. If that word is located in a line, then the algorithm concatenates that line and the following line and then searches that combined line for the entire key phrase. In this way the filtering mechanism can correctly locate key phrases that are extended over two lines. Figure 6.2 shows a sample article as displayed by the information filter.

All of these issues surfaced during the implementation phase. The Access FindFirst and FindNext commands, while providing the basis for the filtering mechanism, were not sufficient and required a great deal of support through Access Basic code.

The Conventional Profile

The conventional profile provides the standard to which the holistic profile is compared. The conventional profile serves as a model of those user profiles incorporated into existing filtering systems, and must be as representative of those as possible. Existing profiles seldom provide user guidance, instead depending on the user's ability to accurately compile a general list of interests.

The implementation of the conventional profile was initiated by the design of input screens, or forms, as they are referred to in Access. Among the implementation issues surrounding the input forms are:

- multiple entries per input field;
- detection of duplicate entries;

- form snapshots; and
- form labels.

Multiple Entries Per Input Field

While the prototype design specifies that the user must be allowed to enter multiple keywords per input field, Access expects an input field to contain a single entry. Therefore a routine was required to extract individual keywords from a comma-delimited list of keywords in an input field. When the user completes data entry in a field and exits the field, the category and keywords are saved to a temporary file. If the input field contains a list of keywords separated by commas, then the string is parsed and each keyword is extracted and saved. Because the keywords are not permanently saved until the save command is issued, the user can return to a previously completed field and add, modify, or delete entries. Figure 6.3 shows a sample input screen containing multiple entries.

Detection of Duplicate Entries

It is not necessary to store duplicate keywords in the profiles, and therefore when the save command is issued the routine checks each keyword before it is saved to be sure that it does not already appear in the profile. If it is already included in the profile then the user is alerted and that keyword is not duplicated in the profile. Similarly, when the user is completing an input form, the input routine prevents the entry of duplicate keywords on a form. For example, if the user enters "construction" in one field, any attempt to enter it in another field will result in a warning message. Thus, duplicates are prevented both on the input forms and in the profiles.

Form Snapshots

Another implementation issue results from the inclusion of a clear button. The clear button is used to clear all of the input fields on a form. Because it is likely that at some point the user will hit the clear button in error, an undo button must also be included to restore the screen to its former state. For this to be possible,

however, the clear routine requires a mechanism to take a snapshot of the screen before the fields are cleared. The phrase "take a snapshot" means that each keyword category and its associated keywords are stored in a table, creating, in effect, a snapshot of the screen. Thus, when the undo button is pressed the snapshot is used to restore the previous values to each of the input fields.

Form Labels

The final implementation issue that had to be dealt with regarding input forms for the conventional profile pertains to the provision for user-defined categories. As explained earlier, user-defined categories are available for those keywords that do not fit into any of the provided categories. In many cases the user may have a category that is unique to his or her particular situation. Because none of the provided category labels adequately describe this particular type of keyword the user may wish to relabel the category, replacing the label "User-defined category" with a more descriptive label. Because of cases such as this, the form labels are not linked directly to the form, because to do so would make them impervious to user updates. Rather, the labels for each form are stored in a table, and copied to the form when it is initially opened, and copied from the form when it is closed. In this way, any changes that the user makes to the user-defined labels are saved so that they will appear the next time the form is opened.

The Holistic Profile

The holistic profile serves as the cornerstone of this research. The holistic profile must encompass several areas of user interests, and must also guide the user through the process of creating a profile.

While the implementation issues related to the conventional profile also impact the holistic profile, additional issues must also be resolved. These issues stem primarily from the incorporation of the template and synonym features into the holistic profile. While these features will be discussed individually, they can no longer be considered separately from the holistic profile. The implementation of the

holistic profile must take these features into consideration so that the overall system will function smoothly.

The first of the implementation considerations is the presentation of templates when the holistic profile input screens are displayed. In addition, the user must be allowed to not only select entries from those templates, but must also be allowed to enter his or her own keywords as well. Finally, synonym lists must be offered to the user during profile input. Therefore, the issues that will be discussed include:

- incorporation of templates;
- keyword entry;
- keyword expansion; and
- form selection.

Incorporation of Templates

The majority of the data entry forms associated with the holistic profile provide templates, and therefore require enhancements to the standard input form developed for the conventional profile. The template feature is designed to suggest possible keywords to the user during profile input. For example, when an accountant opens the ongoing project input form and selects the project type field, the template should provide such project types as "audit," "review," and "tax." Ideally, the template offerings should remain hidden until the user calls for them. Fortunately, Access provides a field type that will accommodate this. An Access combo box is an input field that allows the user to either type in a value or select a value from a predefined list. Its appearance is similar to a normal input field, except that a small down arrow appears to the right of the field. When the user clicks on that arrow a list of predefined choices appears. By using this mechanism, and filling the list with template items for that particular field and domain, the template feature can be implemented. Figure 6.4 shows a sample input screen and an associated template. The difficulty in supplying the correct set of templates to the combo box will be discussed in the following section.

Keyword Entry

One idiosyncrasy associated with combo boxes presents particular difficulties. A combo box is divided into two sections. The first is the list box, which is the input field that always appears on the screen. The second is the combo list, which appears only when the user clicks the down arrow. As stated, a combo box allows the user to either type an entry or select a value from the combo list. However, if the user types an entry and then selects a value from the combo list, the selected value will replace the typed entry. The prototype design requires provisions for any combination of typed entries and values selected from the template list, but the Access implementation of combo boxes does not permit this. However, by "echoing" or duplicating the value from the list box portion of the combo box either in a table or in a hidden field on the form, the holistic profile prototype gives the illusion that the combo boxes are behaving as required. Whenever the user types an entry or selects a value from the combo list, that value not only appears in the list box, but it is also immediately echoed in the hidden field. Then, when the user either types additional keywords or selects values from the combo list, those replace the previous entry in the list box, but at the same time are appended to the echoed entry in the hidden field. This concatenated field, which appears as it should, is then copied from the hidden field back to the list box. So while Access replaces the value in the list box with each new entry, the prototype system provides an *expanding* combo box that makes it appear as if the new entry is actually added to previous entries in the list box portion. The prototype code even inserts a comma between each entry whether the user remembers to or not. In this way the user can perform any combination of entering keywords or selecting values from a template as specified in the requirements. Figure 6.4 shows an input field that combines typed entries and template selections.

Keyword Expansion

Another design requirement calls for the system to provide a keyword expansion mechanism. When a user types in a keyword for which an alternative is available, the user should be notified of the fact and allowed to make a selection from the synonym list. This feature was implemented in the holistic profile prototype as part of the save routine. As noted earlier, the prototype does not permanently save the user's entries every time an input field is completed. The user must have the option of returning to an input field and making any modifications desired. Thus, the entries are saved in a temporary file until the user invokes the save routine. At that point each keyword entry in the temporary file is compared against a synonym list to determine if synonyms are available. If none are available then the keyword is simply saved. If, however, a synonym set is available a synonym screen will appear. This form lists the keyword for which the synonyms are being provided, and provides a combo box with the predefined list filled with the synonyms for that keyword. Because the user must be allowed to select multiple synonyms from the list, the prototype again provides an expanding combo box that grows with each user entry. When all synonyms are selected and the user clicks on the save button, the system asks the user if the new selections are intended to replace or supplement the original keyword. If the user selects "replace" then the system discards the original keyword and stores the selected synonyms in the holistic profile. If the user selects "supplement" then the original keyword and the selected synonyms are stored in the profile.

Form Selection

One additional feature related to the implementation of the holistic profile remains to be discussed. In some cases a user may opt to bypass one of the profile segments. For example, not all industries serve their clients in an information-gathering capacity, so users in those industries may prefer to omit the client type profile. The user can move from input form to input form either forward or backward by clicking the right arrow or left arrow buttons. If no entries have been

made on the form, clicking the right arrow button will bypass that form and move on to the next form. However, if entries have been made and the user attempts to leave the form without first saving the entries, the user will be reminded of the oversight and given the opportunity to save the entries before moving on. Similarly, if the user wishes to make additional entries in a form after saving, there is no need to exit the form. The save routine clears the form so the user can enter additional keywords and then save the form when data entry has been completed.

In summary, the implementation of the holistic profile builds upon the implementation of the conventional profile. However, because the holistic profile offers more features than the conventional profile, enhancements had to be made before the design specifications could be met.

Template Module

The template feature is included to assist the user in profile development by providing a list of potential keywords for the user's consideration. The primary concern when implementing the template feature is how to ascertain which set of templates should be provided. The determining factors are the user's current domain, the current form, and the current input field on the form. Implementation issues revolve around the following:

- determining the current domain;
- determining the current form and field;
- selecting the template; and
- maintaining the templates.

Determining the Current Domain

The user's current domain is obtained from the user's responses on the overview screen. As noted in the previous chapter, the overview screen precedes the input forms and requires the user to enter his or her job title, project type, industry type, and client type. Each of these responses is stored in a table and provides the correct domain when the associated input screen becomes active. For instance, when the functional area input screen is active the domain can be obtained

by extracting the user's entry for job title from the overview table. If the user entered "accountant" for the job title, then the domain for the functional area input form is accountant.

Determining the Current Form and Field

The determination of the current form can be accomplished through an Access form property that identifies the form that is currently displayed. Likewise, the current input field can be obtained by examining the label to the left of the active field.

Selecting the Template

The templates are stored in such a way that by using these three values--domain, form name, and input field--the template module can determine which template set is called for. The templates for each form are stored in individual databases. There is a "title" database that provides the templates for the functional area input form, a "project" database that contains the templates for the ongoing project input form, and an "industry" database that holds the templates for both the organizational environment input form and the client environment input form. By determining the current form name as described above, the correct database can be selected and opened. Separate tables of template values are provided for each input field on the form, with the exception of the user-defined fields. By determining which field is active, the correct template table can be selected and opened. Each table contains a list of template values that are paired with a particular domain. For example, in the industry database there is a technical skills template table that contains a set of templates for an accountant, a construction executive, and a software engineer. When the software engineer enters keywords in the technical skills field, clicking the down arrow provides a template made up of a set of technical skills that may be required of a software engineer. The user can choose one or more of the template values in addition to providing his or her own set of keywords. Because the domains of the users who would be testing the system were known in advance, templates could be provided for each possible domain.

Maintaining the Templates

A template maintenance option is also available. Using this feature the user can not only generate new sets of templates, but can also modify, add, and delete entries in existing templates.

Synonym Module

The synonym module provides a form of profile expansion by pointing out alternative keywords to the user. As in the template module, the primary implementation issue concerning the synonym module is the determination of which synonym set to provide. Similar to the way in which templates are selected, the synonyms are selected based on the current form and the domain. Issues of concern include:

- determining the correct database;
- determining the correct synonym set; and
- maintaining the synonym sets.

Determining the Correct Database

When the user initiates the save routine after completing the entries for a particular form, the synonym feature is activated. If any of the keywords entered in the form have synonyms or alternatives, then the keyword expansion mechanism points them out to the user for consideration. It is important that the correct synonym set is available during this stage. As explained earlier, the name of the form that is currently displayed is available through one of the properties of the form. The synonym feature uses the form name to determine which database to open, again choosing from the title database, the project database, or the industry database.

Determining the Correct Synonym Set

Once the correct database has been opened, the synonym feature uses the current domain to select the correct synonym file. As before, the domain associated

with the current form can be determined by examining the overview table, which contains domain values for job title, project type, industry type, and client type. If a set of synonyms exists for a particular domain, then the database will contain a synonym table with the same name as the domain. For example, the prototype's title database contains synonym sets for a public accountant, a construction executive, and a software engineer. The fact that it was known in advance what types of domains would be used to test the prototype system made it possible to insure that synonyms sets were available for each possible domain. Figure 6.5 provides an example of a synonym offering.

Maintaining the Synonym Sets

The user has the option of generating new synonym sets if additional domains are introduced. In addition, the user can utilize the synonym maintenance option to modify, add, and delete entries in existing synonym sets.

Passive Learning Module

The passive learning module is designed to track keyword performance and, when invoked, alert the user to keywords that are performing poorly. The passive learning module is comparatively straightforward, and as a consequence relatively few difficulties were encountered during its implementation. Those implementation issues include:

- performance variables; and
- performance evaluation.

Performance Variables

The prototype design stipulates that certain values must be stored in order for the passive learning module to perform the required calculations. These values include the number of documents retained due to the keyword, the total score achieved by the keyword based on the user rating of the retained documents, and the number of documents retained by the keyword that have been read and evaluated. The number of documents retained due to the keyword, i.e., the number

of hits, is determined during the filtering process. When the filter locates a keyword in an article the number of hits for that keyword is incremented. The total score and number of documents read are obtained from the user's evaluation of each individual article. When the user initiates the article evaluation, a counter of the number of documents read is incremented for each keyword responsible for the article's retention. When the user assigns a rating to the article that rating is added to the total score accumulator for each responsible keyword.

Performance Evaluation

These variables are used to determine how well a keyword is performing. The passive learning mechanism must be capable of alerting the user to two conditions. The first condition arises when a keyword retains only a small number of documents. The second arises when the documents retained by a keyword are consistently poor. In order to determine what is "few" and what is "poor," performance thresholds must be established. Therefore, performance evaluation must consider the following:

- retaining few documents;
- retaining poor documents;
- performance thresholds; and
- performance evaluation process.

Retaining Few Documents

The first condition arises when a keyword is responsible for the retention of only a few documents. In some cases this is a sign that the keyword is unnecessary and should be omitted. However, it may also be the case that while the keyword is significant, the required information is available only infrequently, in which case the keyword must remain in the profile. An example of such a keyword is "tax act," which is intended to retain information about various Internal Revenue Service tax acts that are passed by Congress on an irregular basis. While this keyword may retrieve information only once a year or less frequently, it is necessary that it remain in an accountant's profile, even though it is seldom used. In either case the

frequency of use of a keyword corresponds to the number of hits variable. If that value is below a certain threshold, then the user is alerted to the fact that the keyword has been responsible for the retention of only *n* documents, and is asked if it should be removed from the system. In this way the user is allowed to decide the fate of a keyword after being alerted to its poor performance.

Retaining Poor Documents

The second condition arises when a keyword is responsible for consistently retaining documents that are of no interest to the user. This is determined by calculating an average score for the keyword by dividing its total score by the number of documents read. A low average score indicates that the keyword should be removed or should be modified to provide greater context so that unwanted stories are no longer retained. If the score is below a certain threshold the user is alerted and given the option to delete the keyword. This allows the user to decide whether to delete or modify the keyword, rather than having the system arbitrarily alter the profile.

Performance Thresholds

The previous paragraphs noted that the passive learning mechanism compares both the number of hits and the average score to threshold values. These thresholds are set somewhat arbitrarily. The average score threshold is set at 2.00, which is the article rating that corresponds to the choice "somewhat useful." If a keyword has an average of somewhat useful or not useful, then it is considered to be a poor performer. This threshold was set at 2.00 since a higher rating indicates that most articles associated with the keyword were considered useful. The number of hits threshold was also arbitrarily set. Because the prototype testing was set for a two week period, the threshold was set at a rather low 5.00. In other words, if a keyword fails to retain more than five documents then the user is alerted. This feature does not lend itself to thorough testing in such a short test period. In the case of the prototype design and testing the threshold had to be rather low because of the limited testing period.

Performance Evaluation Process

When the adapt profile option is selected, the passive learning module cycles through every keyword in the holistic profile, comparing the number of hits and the average score to threshold values. Each questionable keyword is pointed out to the user for a decision on its disposition. In this way the adaptation process is passive rather than automatic, which provides greater control for the user.

Summary

As noted earlier, the implementation of the holistic profile prototype was guided by the prototype design, and constrained by the capabilities and/or limitations of Microsoft Access. In many cases Access provided ideal features for the implementation of the prototype system. This chapter, however, highlighted the shortcomings of Access that necessitated some inventive programming in order to implement a prototype that follows the design. Overall, the selection of Access as an implementation tool was a sound decision, and it resulted in a fully functional prototype system that allows an informative comparison between the two profile types.

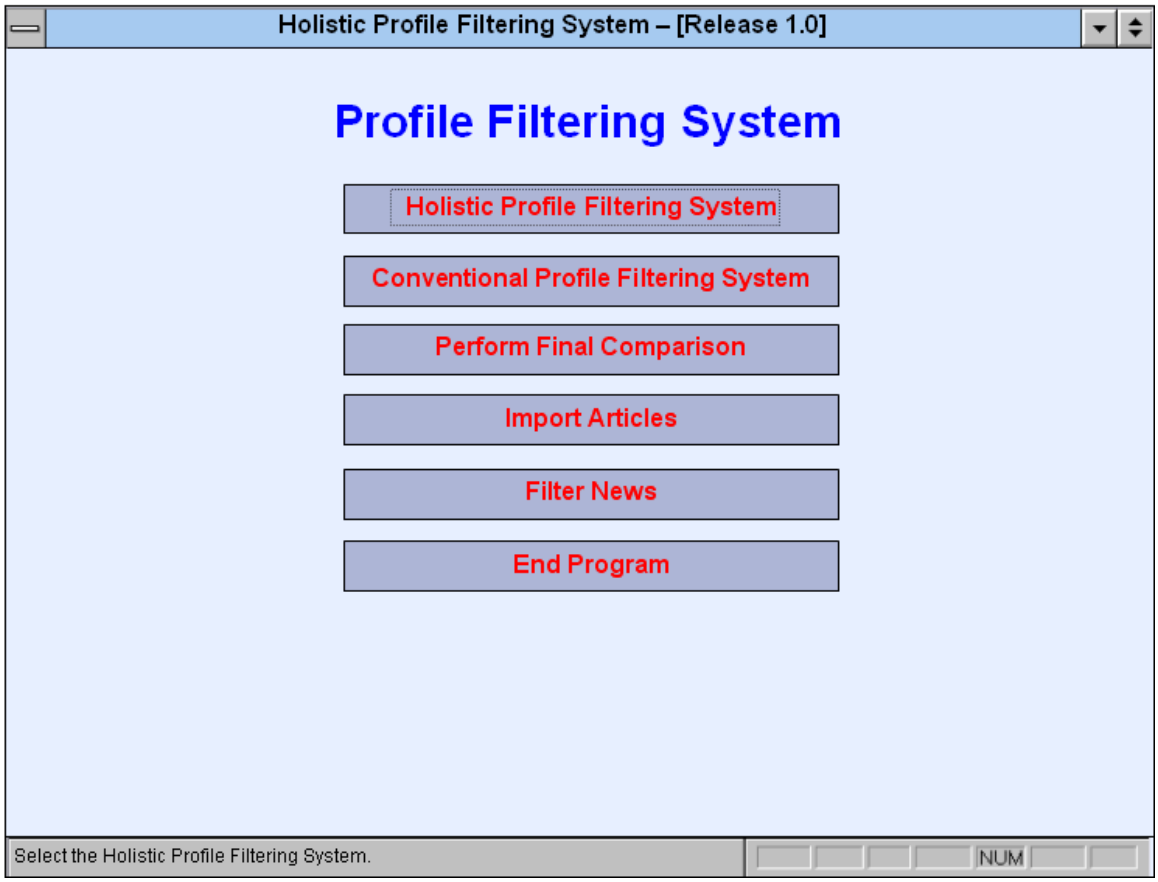

Figure 6.1. Prototype Main Menu.

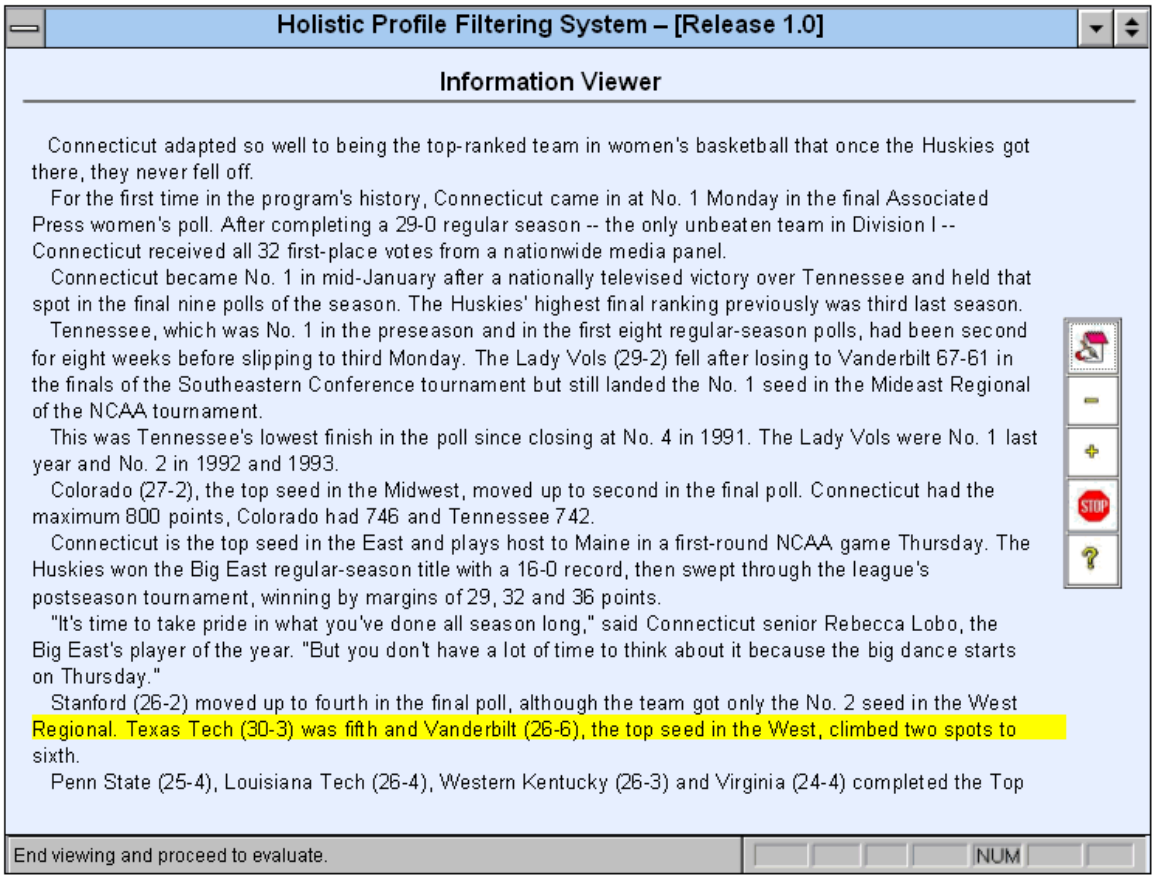

Figure 6.2. Sample Article Displayed by the Prototype.

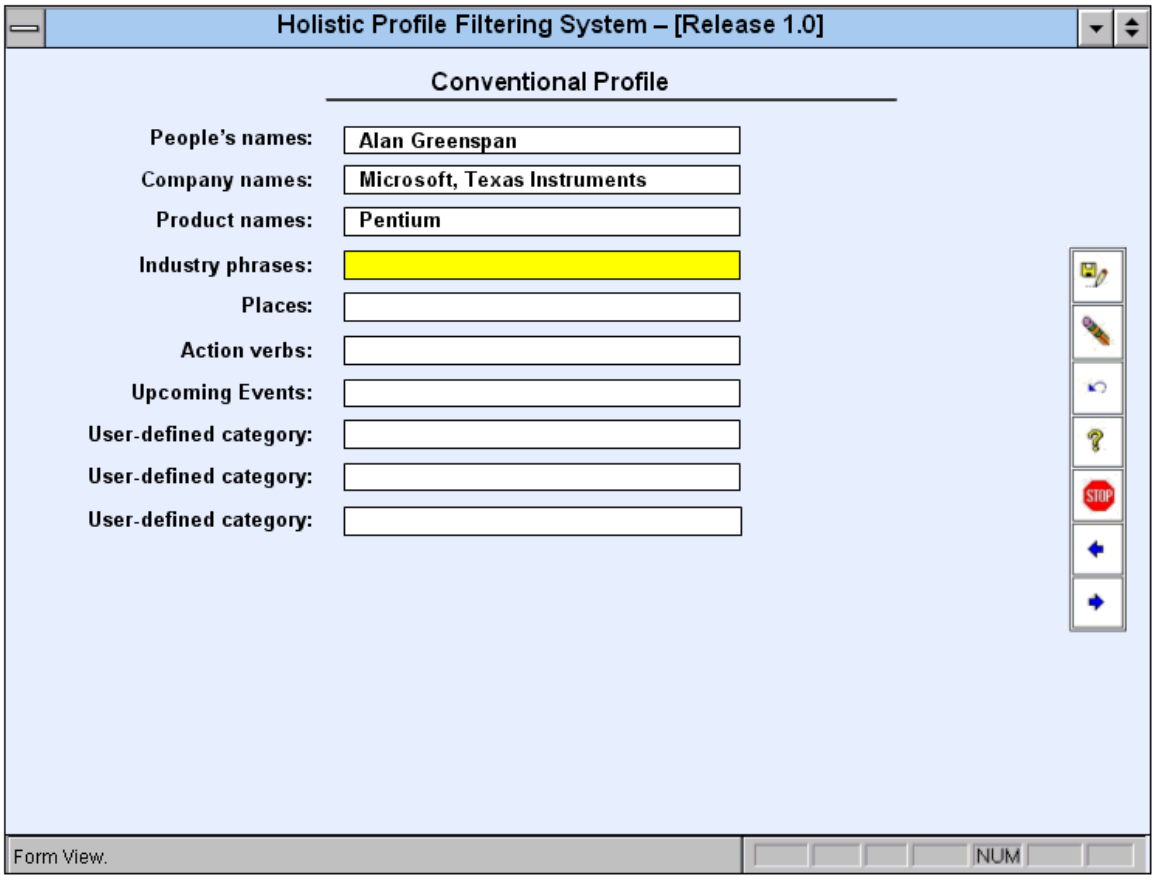

Figure 6.3. Conventional Profile Sample Input Screen.

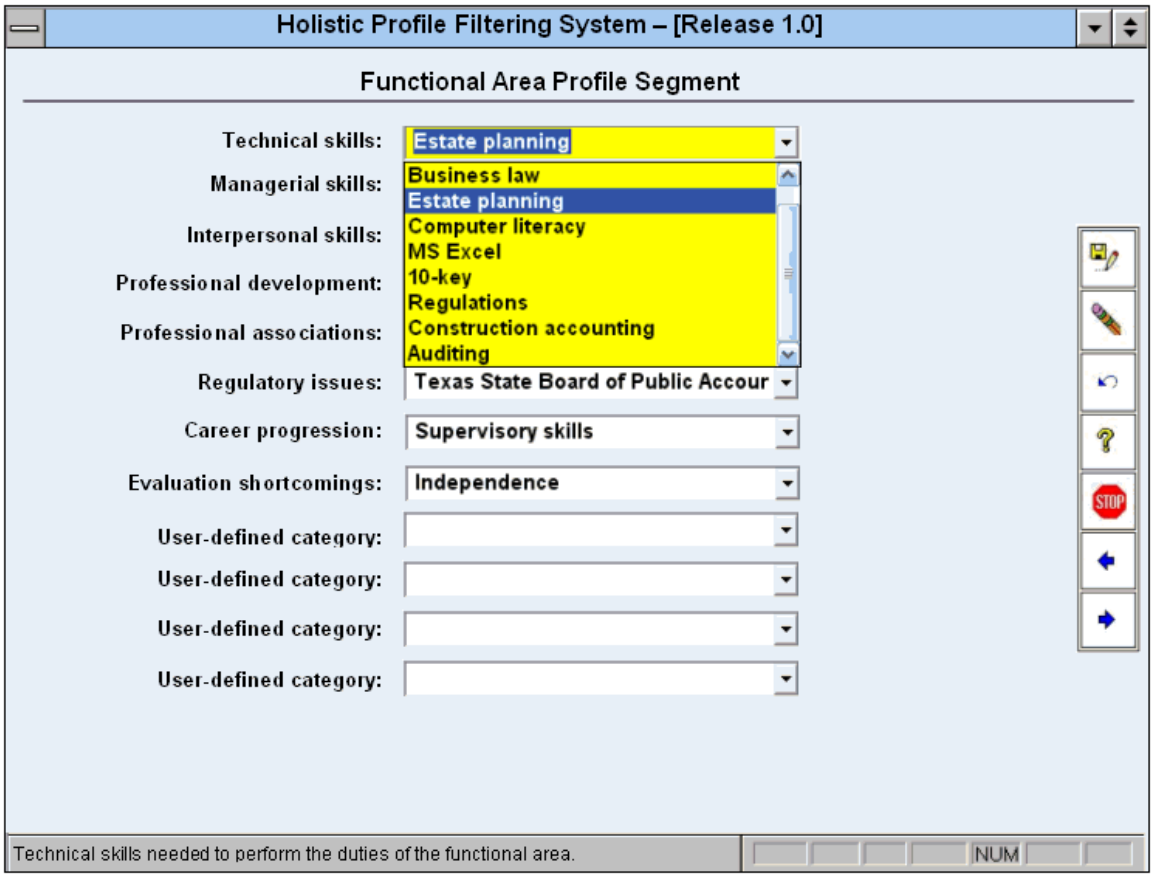

Figure 6.4. Holistic Profile Input Screen.

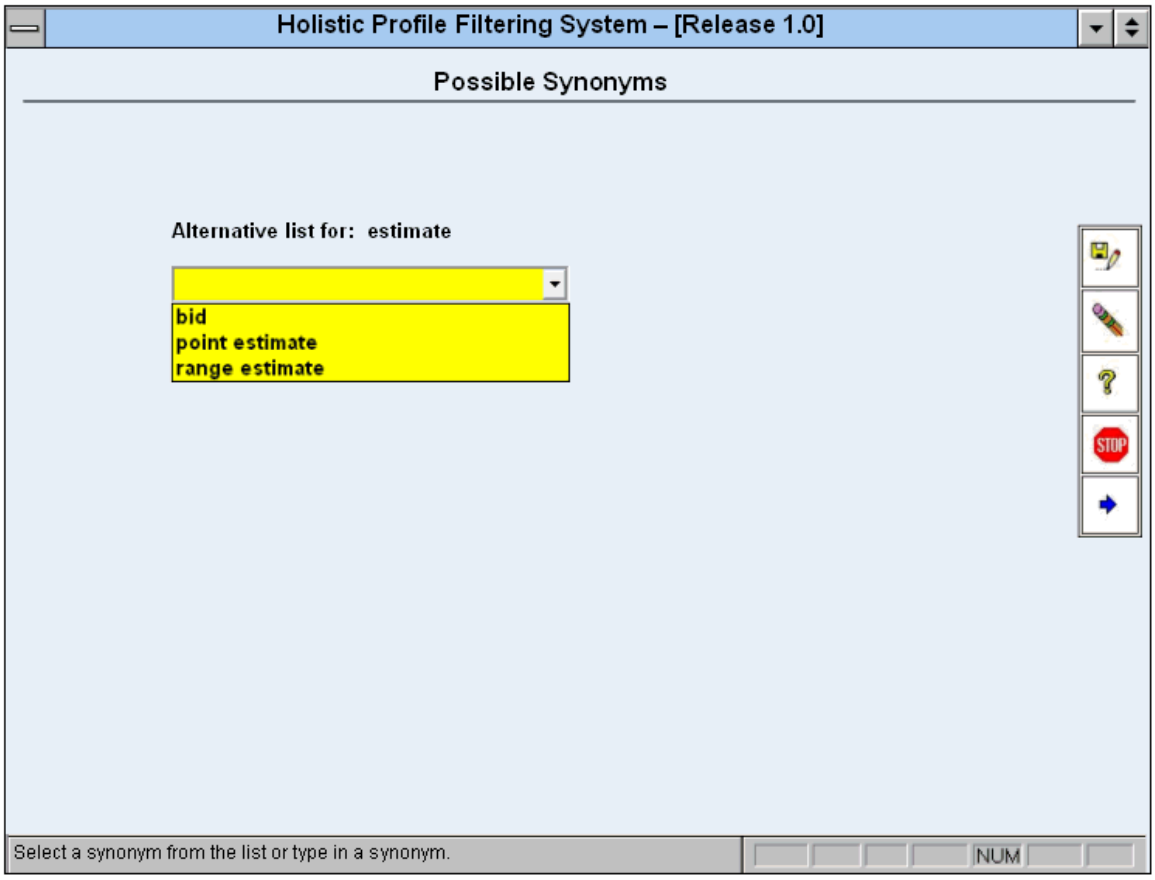

Figure 6.5. Example of Synonym Feature.

CHAPTER VII

VALIDATION OF THE HOLISTIC PROFILE FILTERING SYSTEM

Introduction

Sound scientific practice requires the validation of the methods or techniques by which research is conducted. The term "validation" means to confirm with facts or authority. In the terminology of logic, the purpose of validation is to affirm or deny the truth of a given theory or statement based on a set of known principles or truths.

The conceptual model that was produced by this research was validated by the construction of a prototype system. A prototype provides a model, or a representation of an object, system, or idea in some form other than that of the entity itself [SHAN75]. As such, its purpose is to assist in explaining, understanding, or observing a system.

Prototype Validation

Prototype construction provides a systematic, explicit, and efficient method for researchers to focus their judgment and intuition [SHAN75]. However, it is not enough to simply build the prototype. A prototype system alone is not sufficient to prove a theory [COHE88, COHE89] because the prototype is simply a tool for observation. "A theory cannot be proved from observation (i.e., an antecedent cannot be proved from a consequent)" [BALD94, p. 18], and thus additional measures must be taken.

Prototype validation is a technique in which the results of project development research are evaluated and substantiated:

> To evaluate a model means to develop an acceptable level of confidence that inferences drawn from the performance of the model are correct and applicable to the real world system. The concept of validation should be considered one of degree and not one of an either-or notion; it is not a binary decision variable where the model is valid or invalid. [SHAN75, p. 208]

Application to MIS

Prototype design and consequent model validation are best suited to MIS research focusing on developmental projects. Khazanchi writes that

> From a philosophical perspective, the prototype computer system is a way of confirming (or substantiating) that our model of reality--whatever that may be--is tenable. The prototype system is a tool--akin to the experimental design in empirical research--that provides a means for validating theories or conceptual models. [KHAZ91, p. 40]

The use of prototyping to substantiate conceptual models is a common practice in many applied sciences, such as software engineering, a referent discipline for MIS. O'Leary notes that

> ...prototyping often is used to demonstrate a "proof of concept." Researcher claims for various problem solving approaches must be substantiated by developing a prototype of the system. [OLEA88, p.26]

The validation of prototypes used in MIS research often takes the form of comparison studies. Khazanchi provides an excellent discussion of prototype validation:

> The best way of evaluating "success" is to manually analyze a sample problem and then use the prototype to do the same. The degree to which these analyses are comparable can be taken as the measure for "success" in terms of the verification question. The ability of the prototype system to objectively demonstrate all (or a majority) of the functions specified...can be taken as a measure of "success" in terms of the validation question. [KHAZ91, pp. 163-165]

Validation of the Conceptual Model

While the development and implementation of the conceptual model is a top-down process, the validation of the model is performed in a bottom-up fashion. In the development phase the given requirements are used as a basis for the knowledge-level concepts. The knowledge-level concepts provide the basis for the symbol-level design, on which the prototype design is based. The prototype design is used to guide the implementation of the system, and that implementation is subject to the testing process that generates the test results. In the validation phase

the accuracy of the test results serves to validate the implementation. Successful validation of the implementation validates the prototype design, and the satisfactory validation of the prototype design validates the symbol-level design. Validation of the symbol-level design validates the knowledge-level concepts as well as the overall proposed solution. In a transition, i.e., by implication, this validation process provides a validation of the overall conceptual development process. This process, with respect to the holistic profile, is shown in Figure 7.1.

Validation of the Holistic Profile

A description of comparison studies is provided in Cohen and Howe's [COHE88, COHE89] discussion of various methods of model validation. In the basic form of comparison studies one or more measures of a system's performance are selected, and then both the system and a standard are given the same input and the results are compared on the measures. The validation of the holistic profile prototype was performed through a comparison study using a conventional user profile as the standard against which the holistic profile was compared. The primary assertion that was tested is that there is no difference in the performance of information filtering systems that utilize conventional user profiles and those that make use of the holistic profile. The premise was that if the assertion was rejected, then that provided evidence that one type of profile enhanced the information filtering process. The comparison study was conducted by first filtering an information stream through an information filtering system fitted with a conventional user profile, and then filtering the same stream through the same information filtering system utilizing the holistic profile.

In order to properly describe how the holistic profile system was validated, it is necessary to specify the type of information source, the type of information filtering system, and the problem domain for the profile itself.

Various options for the news source were considered, including various newswires or Usenet news. Requirements included that the news source was text-based and that it provided current news and events. Examples of newswires include

The Associated Press, Comtex, Dow Jones, Federal News Service, Knight-Ridder, and Reuters. These are available through various subscription services, or in some cases over the Internet. Usenet news, or NetNews, is a communication network available via the Internet and consists of thousands of newsgroups that include current news and events [FARR93, NOTE93]. The final decision was to utilize one of the newswire services that are accessible from one of the commercial on-line services.

The information filtering system was developed specifically for the validation phase. Alternatives included the use of commercial news filtering services such as NewsEDGE that accept a user profile and monitor multiple newswires such as those listed above. However, such systems were prohibitively expensive and could not provide a seamless integration of the user profiles being tested. Development of a generic filtering system, while complex, resulted in a system that allowed both the holistic profile and the conventional profile to be "plugged in" and tested independently without any restructuring of the profiles.

The problem domains for the comparison test included public accounting, utility construction management, and semiconductor manufacturing software development. The inclusion of public accounting insured that at least one of the members of the problem domain provided client services in the form of information gathering, and would make use of the client environment segment of the holistic profile. Two individuals from local businesses represented each field for a total of six participants in the comparison study. One participant was unable to complete the study because of other demands on his time. The user's role in the comparison test is described in Figure 7.2. In order to conduct the comparison study each participant in the study compiled a conventional profile of user interests, receiving very little guidance. The conventional profile was based on the profile in NewsEDGE, which provides tips for profile creation but little actual guidance (see Figure 7.3). Similarly, the prototype holistic profiling system was used to guide the construction of a holistic profile for each participant. Over a two week testing period the daily newswire data was processed by the information filtering system using both the conventional profile and the holistic profile. Figure 7.4 provides a

diagram of the process, which resulted in retained information sets 1 and 2. Each participant was required to read both the retained information set and the bypassed information set, evaluating each article with regard to how well it pertained to his or her set of interests. These evaluations were recorded for each article, for each keyword, and for each profile, allowing the system to compile a performance record for each element in the system.

At the end of the testing period the performance record was used to evaluate each profile in terms of two measures referred to as *recall* and *precision*. *Recall* is the proportion of available relevant documents actually retained by the filter, and *precision* is the proportion of the total retained documents that are actually relevant to the user's expressed interests [MAUL91]. These measures are widely used to assess the effectiveness of information retrieval systems [MYAE90], and are equally applicable to the evaluation of information filtering. Recall was assessed by requiring that each participant read and evaluate both the retained news and the bypassed news. In that way the system could record how many relevant stories *should* have been retained by the filter, as well as how many stories actually *were* retained by the filter. For instance, if over the two week period the newswire contained one hundred and fifty stories that were judged by the user to be of interest to him or her, and the holistic profile retained one hundred and twenty of those as opposed to only sixty retained by the conventional profile, the recall score for the holistic profile would be 80% while the conventional user profile managed only 40%. Similarly, the user evaluations can be used to determine the proportion of the retained stories in which the user is actually interested. If the holistic profile retained one hundred and twenty stories of which only ninety were of interest, then the precision score would be 75%. Validation of the system required that the holistic profile attain a higher score in recall than the conventional profile, and an equivalent or greater score in precision for a majority of participants.

However, only after the testing period was complete and the analysis of the results was underway did it become apparent that looking at recall and precision as independent measures does not give a complete picture of a system's performance.

An information filtering system with high recall but low precision retrieves a high proportion of the available interesting documents, but also allows a great deal of unwanted material through the filter. On the other hand, a system that achieves high precision but low recall allows little unwanted material through the filter, but is so selective that it misses a great deal of material that should have been retrieved. Therefore, this study introduced a new measure for information filtering and information retrieval, referred to as *efficiency*. Efficiency is derived from the interaction between recall and precision. It takes into account the impact of low precision on high recall, or conversely the impact of low recall on high precision. In either case, it allows the evaluator of the system to readily see the interaction between recall and precision and to more accurately measure the system's performance. As with recall, validation of the system required that the holistic profile attain a higher score in efficiency than the conventional profile.

Validity Issues

Prototype validation must also consider validity issues. Specifically, both internal validity and external validity are of particular concern with the use of prototypes.

Internal validity concerns the allocation of variance and the proper use of the selected tool [COOK79]. In many prototypes, a model is developed and compared to a standard. Certain criteria are established, and if the prototype meets or exceeds these criteria, then the proposed system is an improvement over either the standard or an existing system. The internal validity, or reliability, of the comparison is dependent on the criteria or standard selected. In the case of the holistic profile, the standard against which the model was compared was the conventional profile. As noted above, the comparison was performed in terms of recall, precision, and efficiency. Selection of the most appropriate standard available, as well as commonly accepted measures, helps to assure the internal validity of the holistic profile prototype.

External validity refers to how appropriate the results of a study are in the real world [COOK79]. Because a prototype is by definition not a fully operational system, it may not incorporate all aspects of the proposed conceptual model. Thus, while the prototype should be sufficiently generalizable, there are limitations on external validity. The holistic profile system has been designed to be generalizable across organizations, departments, and positions. This generalizability was demonstrated by the application of the holistic profile to different problem domains, e.g., public accounting, construction management, and software development. These domains are sufficiently different to show generalizability across organizations and positions. In addition, the holistic profile, while tested with a generic filtering system, was designed to be applicable to any type of filtering system that accepts a pre-defined profile. Further, the inclusion of user-defined categories in each segment of the holistic profile helps to insure that the holistic profile is generalizable across users.

Determining the Results

As discussed earlier, the primary measures by which the success of the holistic profile can be measured include recall, precision, and efficiency. Recall was defined as the percentage of the relevant documents contained in the information stream that are actually retained by the profile, precision is the percentage of all documents retained by the filter that are actually of interest to the user, and efficiency is the interaction between recall and precision.

In order to arrive at a result for each of these measures the system utilizes the user evaluations of each article to track several variables. The number of articles retained (N_{HIT}) reflects the total number of articles actually retained by the system, including both relevant and irrelevant documents. The number of articles rejected (N_{BAD}) refers to the number of documents that were retained by the filtering system but were nevertheless rated by the user as being of no interest to him or her. The number of articles missed (N_{MISS}) is a counter of those articles that

were not retained by the filtering system, but were rated by the user as being of interest.

The two factors involved in the calculation of recall are the number of relevant documents retained by the profile and the number of relevant documents that are available. The number of relevant documents retained by the profile can be obtained by subtracting the number of rejected documents (N_{BAD}) from the total number of retained documents (N_{HIT}). The number of available relevant documents can be obtained by subtracting the number of rejected documents (N_{BAD}) from the total number of retained documents (N_{HIT}) and adding the number of relevant documents that were not retained by the filter (N_{MISS}). Therefore the equation for recall is

$$Recall = \frac{N_{HIT} - N_{BAD}}{N_{HIT} - N_{BAD} + N_{MISS}}. \qquad (7.1)$$

The two factors involved in the calculation of precision are the number of relevant documents retained by the profile and the total number of documents retained by the profile (N_{HIT}). The number of relevant documents retained by the profile can be obtained by subtracting the number of rejected documents (N_{BAD}) from the total number of retained documents (N_{HIT}). Therefore the equation for precision is

$$Precision = \frac{N_{HIT} - N_{BAD}}{N_{HIT}}. \qquad (7.2)$$

Because efficiency reflects the interaction between recall and precision, it can be thought of as the weighted recall. Efficiency is the reduction in recall caused by a lack of precision, or the proportion of retained documents (N_{HIT}) that were not relevant (N_{BAD}). Therefore, a measure of efficiency can be arrived at by multiplying recall by the weight

$$Weight_{Recall} = 1 - \frac{N_{BAD}}{N_{HIT}}. \qquad (7.3)$$

Efficiency can be calculated using the following equation:

$$Efficiency = Recall * \left(1 - \frac{N_{BAD}}{N_{HIT}}\right). \quad (7.4)$$

Efficiency can also be thought of as a weighted precision. Efficiency is the reduction in precision caused by a lack of recall, or the proportion of relevant documents that were not retained (N_{MISS}). Therefore, efficiency can also be arrived at by multiplying precision by the weight

$$Weight_{Precision} = 1 - \frac{N_{MISS}}{N_{HIT} - N_{BAD} + N_{MISS}}. \quad (7.5)$$

Efficiency can then be calculated using the following equation:

$$Efficiency = Precision * \left(1 - \frac{N_{MISS}}{N_{HIT} - N_{BAD} + N_{MISS}}\right). \quad (7.6)$$

Because equation 7.3 is equivalent to equation 7.2, and equation 7.5 is equivalent to equation 7.1, the equation for efficiency can be expressed as the product of recall and precision.

$$Efficiency = Recall * Precision. \quad (7.7)$$

Evaluation of the Results

Based on equations 7.1, 7.2 and 7.7, recall, precision, and efficiency were determined for both the conventional and the holistic profile. These values can be seen in Table 7.1. By looking at the overall rating for each profile in the final row, it can be seen that the holistic profile achieved a score of 66.23% for recall as opposed to 45.18% for the conventional profile. The holistic profile had a score of 70.09% for precision, while the conventional profile scored only 59.84%. Finally, the holistic profile achieved an efficiency score of 46.42% compared to only 27.03% for the conventional profile. As stated earlier, the validation of the conceptual model required that the holistic profile attain a higher score in both recall and efficiency

than the conventional profile, and an equivalent or greater score in precision. All of these conditions have been met, thus validating the conceptual model. The assertion that there is no difference between the profiles can be rejected, which indicates that the holistic profile is indeed an improvement over the conventional profile.

Analysis of the Results

Table 7.1 reveals that in some cases the holistic profile performed much better than the conventional profile, while in other cases it resulted in much smaller improvements. An analysis of the behavior of the participants in the study helps to explain these discrepancies.

The scores attained by the holistic profile were most impressive for subjects 1 and 2. Each of these subjects created their profiles immediately upon receiving the prototype system and initiated the filtering process the day that the test began. An examination of their profiles indicates that there is approximately a 4:1 ratio between the number of entries in the holistic profile and the number of entries in the conventional profile. These subjects performed the comparison test as directed and devoted careful preparation to their profiles. As a consequence the holistic profile performed far better than its conventional counterpart for these subjects.

The performance measures were much closer for subjects 3 and 4. Like all of the test subjects, subjects 3 and 4 were business professionals, but unexpected demands on their time prevented them from creating their profiles until the test period reached the halfway point. Consequently, their holistic profiles are only slightly larger than their conventional profile, reflecting the subjects' need to get the profiles completed as quickly as possible. In the case of subjects 3 and 4 there is a ratio of approximately 1.5:1 between the size of the holistic profile and the conventional profile. Because these subjects fell behind in the test they were unable to devote enough time to creating a thorough holistic profile, and as a result the holistic profile performed only marginally better than the conventional profile.

Subject 5 had the most unusual results. The recall achieved by both profiles was remarkably low, while the precision was extremely high. This can be attributed to the fact that this individual's profile was overly specific. For example, one profile included the key phrase "Texas Tech University Lady Raiders." While some sports articles may contain the phrase "Texas Tech" and others may include the phrase "Lady Raiders," it is extremely unlikely that any article will contain the entire phrase as specified. Because the profiles were so specific, very few articles were retained, but those that were retained were usually of interest to the subject. Thus, the recall was low while the precision was high.

The variation in the performance measures reveals the influence that the user can have on the effectiveness of any filtering system. Profile creation is a difficult process, even with the assistance provided by the holistic profile, and it requires careful consideration and reflection to develop a comprehensive profile. If the user is unable to take the time to develop an adequate profile, then the performance gains that are made possible by the holistic profile will not be realized.

Understanding the Results

Although the margin of difference varied from subject to subject, the holistic profile consistently performed better than the conventional profile. However, to say that the holistic profile is an improvement over conventional profiles is not enough. The factors that contribute to that improvement must be examined. First and foremost among those factors is the comprehensiveness of the holistic profile. By comparing the holistic and conventional profiles of each of the participants in the comparison study, it was seen that while the size differential between the profiles may not always be appreciable, the holistic profile is consistently broader in scope and more complete. This can be attributed to the fact that the holistic profile provides more guidance to the user during profile creation. Various categories are suggested to make the user more aware of how his or her information gathering can impact not only personal knowledge, but also how it can assist in performing the tasks associated with one's job, in the success of ongoing projects, in the

organization's preparedness for environmental changes, and in the level of service provided to clients. By forcing the user to consider his or her information needs in each of these areas, the holistic profile becomes more comprehensive than conventional profiles.

Another factor in expanding the coverage of the profile is the provision of templates to guide the user and to suggest potential keywords. In many cases the participants relied heavily on the templates, using the profile suggestions as a basis for their initial profile and modifying them as the information filtering process progressed.

Finally, keyword expansion in the form of a synonym feature also helped to make the holistic profile more complete. Useful for expanding acronyms as well as providing less common industry or career phrases, the synonym feature helps to expand the profile and make it more complete by supplementing the user's keyword selections.

All of these features taken together make for a more complete and comprehensive profile, expanding the coverage and accuracy of the information filtering process. It has been claimed that an information filtering system is only as good as the profile that guides it, and the results of the comparison study show that the holistic profile clearly enhances the information filtering process.

Validation of Knowledge-Level Propositions

Aside from their use in the comparison study, the test results that stemmed from the use of the holistic profile also served to validate the knowledge-level propositions. Each proposition was validated on an individual basis by showing that the conditions are met, i.e., the stated knowledge or design is present, and the desired behavior is achieved.

The first knowledge-level proposition states that knowledge of a holistic profile improves the level of recall associated with an information filtering system. As shown in Table 7.1, an information filtering system utilizing a holistic profile has greater recall than a system relying on a conventional profile.

The second knowledge-level proposition asserts that knowledge of a user's interests enables the system to provide information satisfying general interests as well as input to the user's mental model. Examination of the keyword performance table for each participant in the study shows that many of the keywords associated with the personal segment of the holistic profile were responsible for the retention of several articles that the user rated as being of interest. Because any new information helps to shape the user's view, or mental model of the world, the information retained by the personal interests segment of the holistic profile, and indeed all of the other segments as well, contributes to the growth and evolution of the user's mental model.

The third knowledge-level proposition states that knowledge of functional area interests allows the information filtering system to assist the user in increasing his or her professional skills and knowledge. The functional area segment of the holistic profile contained several career-related keywords that perform this function. Further examination of the keyword performance table for each participant in the study shows that many of the keywords associated with the functional area segment of the holistic profile were responsible for the retention of articles that the user rated as being of interest. By allowing the user to keep current on developments in his or her field, the information filtering system assists the user in honing his or her skills and becoming a more productive and better qualified employee.

The fourth knowledge-level proposition makes the claim that knowledge of ongoing projects assists the information filtering system in providing information that may affect the successful outcome of those projects. For the current project segment of the holistic profile each user was asked to provide keywords describing various facets of successful project management. Because of the short duration of the comparison study, it is difficult to claim that the holistic profile was responsible for the successful completion of any of the projects associated with the participants in the study, but the keyword performance table for each participant in the study reveals that project-related keywords did result in the retention of useful articles.

The fifth knowledge-level proposition states that awareness of influential factors in both the organizational environment and the client environment permits the holistic profile and information filtering to be used as an environmental scanning tool. Each of the participants in the study were guided by the holistic profile to include several keywords that would alert them to changes in their environment, including changes in their client base, suppliers, competitors, technological factors, geopolitical factors, etc. One set of participants was also involved in providing client services that often take the form of information gathering, so they employed the client type segment of the holistic profile as well. Although a two week period is too brief to assess the environmental scanning capabilities of the tool, the keyword performance table shows that the holistic profile was responsible for retrieving articles that reflected changes in the organization's environment, and also in the client environment where applicable.

The sixth knowledge-level proposition asserts that knowledge of domain-specific keywords allows the system to provide greater user guidance in developing the profile by providing alternative keywords. This assertion is easily shown by comparing the comprehensiveness and keyword selection in the participants' holistic and conventional profiles. In every case the holistic profiles are considerably more complete than their conventional counterparts, and each holistic profile made extensive use of template-provided keywords.

The final knowledge-level proposition states that the use of performance objects to retain knowledge of not only how many documents were retained by a keyword but also the user's evaluation of those articles allows the system to track keyword performance and alert the user to keywords that are performing poorly. This feature was demonstrated by the prototype filtering system that reminded the user after one week of testing to utilize the adapt profile option. This option was used to examine each keyword in the profile and alert the user to keywords that had either retained few documents, or were associated with the retention of poorly rated documents. While this option requires more than a week to accurately assess

a keyword's performance, it still demonstrated the passive learning capacity of the holistic profile.

Validation of Symbol-Level Propositions

The results of the comparison test can also be used to validate the symbol-level propositions. The symbol-level propositions will be validated on an individual basis by showing that the stated design is present and the desired behavior is achieved.

The first symbol-level principle states that by identifying specific areas of user interests and by describing those areas by a set of keyword categories, the recall and comprehensiveness of the information filtering process can be increased. An examination of the structure of the holistic profile reveals that the profile is indeed segmented into various areas of user interests and that each of those segments is further divided into descriptive keyword categories. The results of the comparison study, as noted above, show that using the holistic profile results in greater recall and comprehensiveness than conventional profiles. Thus, this proposition is validated.

The second symbol-level principle asserts that a system that is capable of tracking keyword performance in terms of usage and relative success can passively learn and adapt by suggesting profile modifications to the user. The adapt profile feature of the holistic profile prototype, which bases its recommendations on the keyword performance knowledge base, provides exactly this capability, and validates this proposition.

A third symbol-level principle states that a modularized structure permits the inclusion of domain-specific templates. Each segment of the holistic profile, as implemented in the prototype, has associated with it domain-specific templates that are intended to assist the user in keyword selection. For example, if the user is in public accounting, when he or she invokes the input screen for the functional area segment of the profile, each keyword category in that segment has associated with it

a template of potential keywords for use by a public accountant. The inclusion of this feature serves to validate this proposition.

Another symbol-level principle states that a modularized structure allows the isolation of segments that are domain dependent, thus allowing them to be duplicated and shared by others. Each segment of the holistic profile, as implemented in the prototype, is created individually and therefore can be duplicated and shared. Again, this principle is validated by the implementation.

The final symbol-level principle requires the development of a domain-specific set of keyword synonyms. This feature is provided by the holistic profile prototype. When the user saves his or her entries for a particular segment, the keyword expansion mechanism scans a domain-specific synonym file for that user's context in that particular career, project, or industry, and provides a set of synonyms for the user's consideration. The inclusion of the keyword expansion mechanism validates this principle.

Thus, by showing that the structure of the holistic profile incorporates the features required by the symbol-level principles, each principle is validated.

Summary

Prototype development and modeling are but one component of a research method, and without proper validation are subject to question. Validation is a vital step in the research process because it helps to establish the acceptability of a model's results, and thus generates confidence that the model's performance accurately reflects the real-world system being modeled.

Further, by validating each of the propositions upon which the research is based, the overall conceptual design is validated. By validating the conceptual development process both through a comparison study and through validation of the underlying propositions, the conceptual development process is further substantiated.

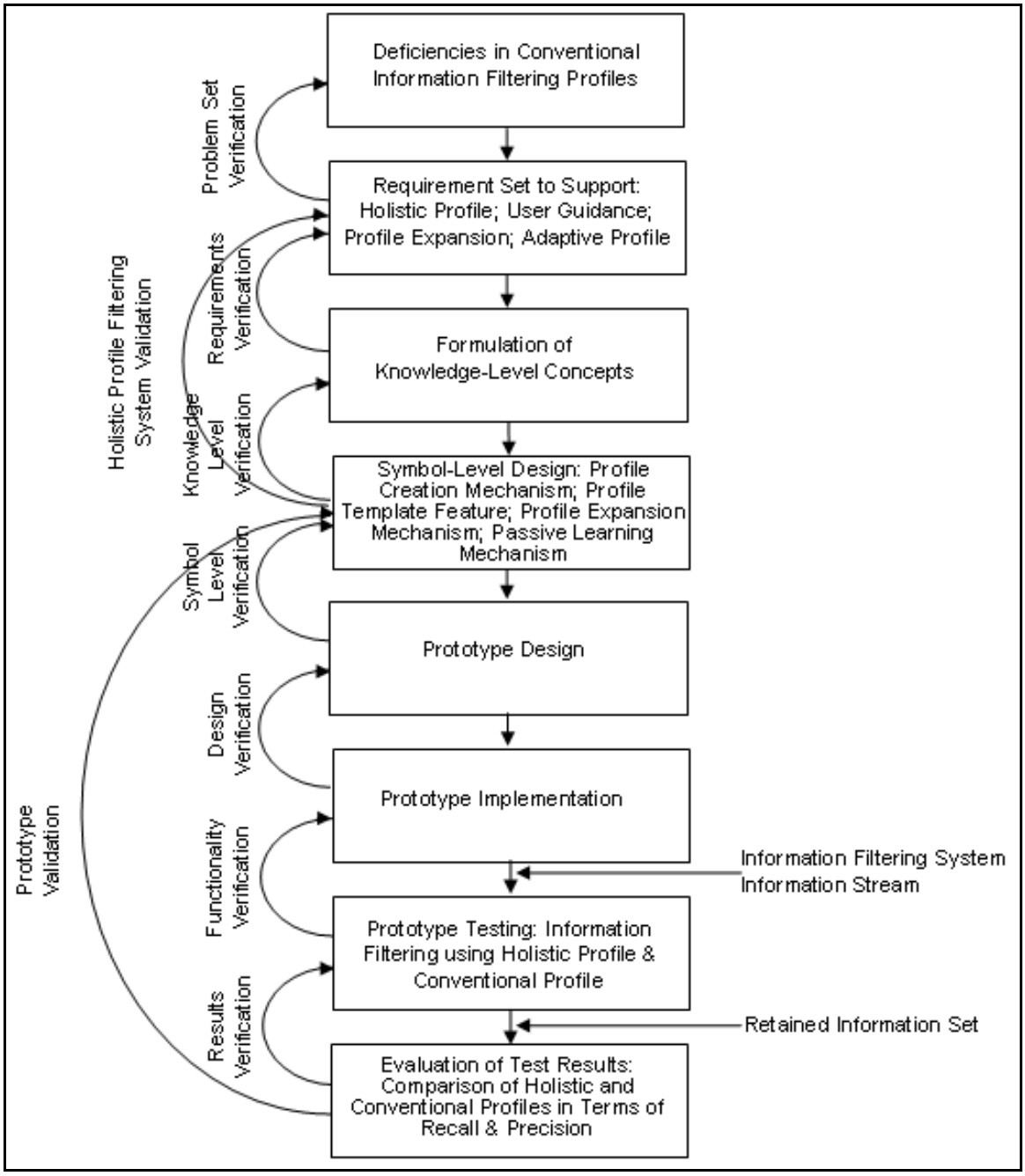

Figure 7.1. Validation Process.

Each user will create both a conventional profile and a holistic profile to describe his or her interests.

Each day for approximately two weeks the daily news from one of the major news wires will be delivered to each test site and will be filtered through both profiles.

Each day the user will read and evaluate both the retained information set as well as the bypassed information set.

At the beginning of the second week the user will perform the Adapt Profile operation of the holistic profile.

At the end of the final week the user will execute the Final Comparison operation which will tabulate the final results.

Figure 7.2. Description of the Comparison Test.

People:
List company officers, company directors. industry leaders, and political figures.

Companies:
Include competitors, customers, suppliers, and investments. Acronyms such as GAAP can be used, but be sure that they have only one meaning. For example, USC is both the University of Southern California and the University of South Carolina.

Products:
List specific products and generic industry names.

Industry Phrases:
Include industry standards and generic industry names.

Locations:
List locations of customers, business associates, suppliers, and stakeholders.

Actions:
Include industry actions such as bid, merge, restructure, and takeover.

Events:
Include significant upcoming events.

Figure 7.3. Guidance Provided for the Conventional Profile.
[Adapted from NewsEDGE]

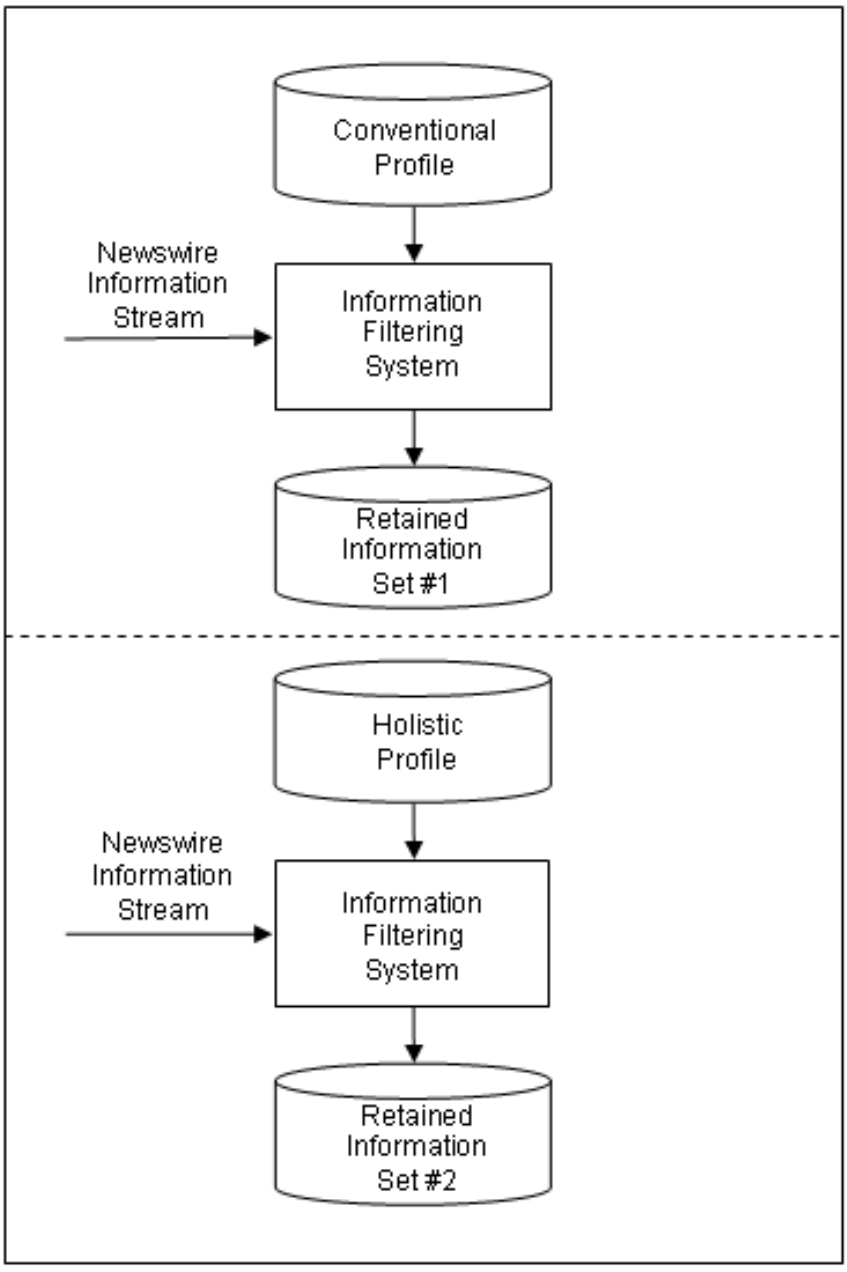

Figure 7.4. Diagram of Comparison Study.

Table 7.1. Results of Comparison Test.

		Profile	
Subject	Measure	Holistic	Conventional
1	Recall	82.20%	22.88%
	Precision	85.84%	71.05%
	Efficiency	70.56%	16.26%
2	Recall	94.24%	42.86%
	Precision	97.06%	41.87%
	Efficiency	91.47%	17.95%
3	Recall	67.92%	54.72%
	Precision	48.00%	44.62%
	Efficiency	32.60%	24.42%
4	Recall	71.88%	70.31%
	Precision	52.87%	41.67%
	Efficiency	38.00%	29.30%
5	Recall	14.89%	2.13%
	Precision	66.67%	100.00%
	Efficiency	9.93%	2.13%
Overall	Recall	66.23%	45.18%
	Precision	70.09%	59.84%
	Efficiency	46.42%	27.04%

CHAPTER VIII

CONTRIBUTIONS, LIMITATIONS, AND GENERALIZATIONS

Summary of the Research

It is generally accepted that information overload is becoming an increasingly serious problem. Information filtering provides a partial solution to the problem of information overload by filtering incoming information and removing data that fail to meet criteria specified by the user. This filtering process, however, is based on a user profile, and if the user profile is either inaccurate or incomplete then the effectiveness of the filtering process is seriously diminished. In short, the quality of an information filtering system is dependent on the quality of the underlying user profile.

The holistic profile was developed in order to rectify this shortcoming in information filtering systems. The holistic profile expands the user profile to encompass not only general concerns, but also concerns associated with the user's profession, with any projects with which the user is currently involved, with the type of industry in which the user is employed, and with the type of clients that the user serves. By expanding the scope of the profile and by suggesting keyword categories to assist the user in formulating the profile, the holistic profile not only expands the description of user interests but also expands the coverage and effectiveness of the overall filtering system.

The notion of a holistic profile and the increase in information filtering effectiveness is substantiated through the implementation and testing of a prototype system. The prototype system demonstrates that, when used properly, the holistic profile results in a significant increase in the recall, precision, and efficiency of the information filtering system.

Contributions of the Research

Research into the holistic profile benefits not only users of information filtering systems, but also the study of information filtering itself.

The contributions to the user are demonstrated by the benefits that accrue from use of the holistic profile. While information filtering as a whole reduces information overload, the holistic profile both increases the scope while increasing the accuracy of the filtering process. Because the holistic profile is designed to encompass a wider range of user interests, more information can be made available to the user. However, the passive learning mechanism reduces the amount of unwanted material that is allowed through the filter, thereby increasing the accuracy of the system. Thus, the holistic profile provides an approach by which an information filtering system can be made to yield greater recall, precision, and efficiency.

The inclusion of the holistic profile in an information filtering system also provides the business world with an environmental scanning tool. If the organizational environment profile is designed with this end in mind, environmental scanning can be performed not only for the corporation, but also the clients that are served by the corporation. Information gained from environmental scanning can assist the organization in the strategic planning process.

This research also has an impact on information filtering as a whole. One contribution to information filtering is that the holistic profile allows filtering to become a true management information systems tool. While the user profile that is incorporated into most information filtering systems consists of little more than general user interests, the various segments of the holistic profile insure that information filtering can help the user to acquire information that impacts the employee, the organization, and the client base. By redirecting the purpose of information filtering, the holistic profile allows information filtering to become a true corporate tool.

Another contribution of the holistic profile to information filtering research is that is redirects the focus of the research from "how" to "what." Previous research into information filtering has focused on how to develop better filtering techniques. Examples include keyword matching, semantic content, and document representations. However, little if any research has been devoted to determining

what should be included in the user profile to make information filtering a more robust tool, which is remarkable considering the fact that the user profile determines the effectiveness of the entire information filtering process. This research will hopefully draw some attention to an area that has been thus far overlooked.

Finally, this research introduces a new measure that is applicable to both information filtering and information retrieval. Efficiency more accurately measures a system's performance by relating the recall and precision measures. The introduction of this measures allows the researcher to evaluate system performance as a whole, rather than considering two seemingly independent variables.

Limitations of the Holistic Profile

Two primary limitations of this research are immediately apparent. The first limitation stems from the type of information filtering mechanism that was chosen to implement the conceptual design. As noted in the previous section, there are a variety of approaches to information filtering, the least effective of which is keyword matching. However, the majority of information filtering systems that are currently available perform some sort of keyword matching, and the results of this research can be applied to make such systems more effective. Although keyword matching may not be as effective as other filtering techniques such as latent semantic indexing, it provides an easily implemented vehicle for demonstrating the validity of the holistic profile.

The second limitation involves the various segments that make up the holistic profile. At present the segments include personal interests, professional interests, project concerns, organizational interests, and client service interests. It is by no means certain that these five segments are sufficient to describe every user's interests. It may be that additional segments are needed to further expand the coverage of the holistic profile. The modular nature of the holistic profile, however, makes the incorporation of additional segments an effortless task.

Generalizations of the Holistic Profile

Although the prototype is fully functional, it is not fully developed, which makes it necessary to generalize the results of this study. The improved recall, precision, and efficiency demonstrated by the prototype system indicate that a fully developed information filtering system using a holistic profile will yield improvements over those currently in use. While the results of holistic profile prototype can be generalized to apply to any filtering system that accepts a keyword-profile, the basic idea behind the holistic profile, i.e., the multiple segments and their applicability to industry, can be applied to other types of filtering systems as well. A simplified example demonstrates the application of the holistic profile to latent semantic indexing.

Latent semantic indexing (LSI) calculates associations among terms and documents with the assumption that there is an underlying structure in the pattern of word usage across documents [FOLT92]. A description of terms, documents, and user queries based on the underlying semantic structure is used for representing and retrieving information [FOLT92]. User interests can be described in a document profile that consists of documents that the user has previously rated as useful [FOLT92].

The ideas behind the holistic profile can be applied to a latent semantic indexing system in the following way. The initial document profile created by the user can be a holistic document profile. A holistic document profile is a set of abstracts developed by the user describing his or her personal interests, functional area interests, current project interests, organizational environment interests, and client interests. Guidelines for these abstracts can be provided in a manner similar to the presentation of keyword categories in the holistic profile prototype. These abstracts can be incorporated into the user's document profile and assigned the highest possible rating. In this way the LSI filtering system can be "trained" to

retain documents that best reflect the user's interests as described in the holistic document profile.

This example serves to demonstrate that the holistic profile is generalizable across other types of filtering systems. While the prototype system utilizes a filtering system that employs keyword matching, the holistic profile is by no means limited to such systems.

Future Work

This research area is still largely unexplored and as a result the potential for future work is extensive. Not only can there be improvement to the current design, but the overall idea can be applied in other areas.

The current design can be enhanced in several ways. The first enhancement involves the addition of logical operators to keyword specifications. The holistic profile prototype currently does not permit the user to specify keywords that are joined by an AND operator. For example, the user might want the filtering system to retain documents that contain both keyword *a* as well as keyword *b*. Similarly, the user might want the filtering system to exclude any documents that contain a certain keyword. In this case the keyword could be preceded by a NOT operator, which would indicate to the filtering system that documents containing that keyword should not be retained. These features will be implemented in the future.

Another feature of the current system is the profile expansion mechanism, or synonym feature. The prototype utilizes a very primitive synonym generator that can be greatly improved. Future research will focus on a more effective technique for keyword expansion, such as thesaurus algorithms, query expansion techniques, and WordNet.

Yet another feature of the current design that will be investigated in future research is the passive learning mechanism. In the current design the adaptation feature is invoked by the user, and only then is an analysis performed on keyword performance. Investigation into intelligent agents and their application to the

passive learning mechanism will reveal the feasibility of making keyword analysis a continuous process in order to improve filtering precision.

Finally, the application of the holistic profile to other types of information filtering systems, as discussed in the previous section, will be investigated more thoroughly. The various types of filtering systems must first be identified, followed by a determination of the applicability of the holistic profile to each system. Only then can prototype systems be developed to further test the usefulness of the holistic profile.

Evaluation of the Holistic Profile

The holistic profile is superior to existing profiles because it assists the user in developing a profile that is more comprehensive and better-suited to his or her needs. The holistic profile structures the profile-generating process and guides the user in specifying keywords for inclusion in the profile. Unlike existing profiles, the holistic profile's coverage of topics extends beyond the range of user interests into the domains of functional area, current project, organizational environment, and client type. Thus, the holistic profile is an improvement over existing profiles not only because it provides an expanded perspective, but also because its very structure provides guidance to the user in the process of profile development. All of these factors make it possible for the information filtering system to achieve greater recall, precision, and efficiency.

REFERENCES

[ACKO67]Ackoff, R.L., "Management Misinformation Systems," Management Science, Vol. 14, No. 4, (December 1967), pp. B147-B156.

[ACKO62]Ackoff, R.L., S.K. Gupta, and Minas, J.S., Scientific Method, John Wiley and Sons, Inc., New York, 1962.

[ADAM89]Adams, R.H., "Management, Analysis and Planning for Skill Development...in the 90's," SAM Advanced Management Journal, Autumn 1989, pp. 34-40.

[AGUI67]Aguilar, F.J., Scanning the Business Environment, Macmillan, New York, 1967.

[ALLE90]Allen, R.B., "User Models: Theory, Method, and Practice," International Journal of Man-Machine Studies, Vol. 32, (1990), pp. 511-543.

[ANDE90]Anderson, J.R., The Adaptive Character of Thought, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1990.

[ANSO75]Ansoff, H.I., "Managing Strategic Surprise by Response to Weak Signals," California Management Review,, Vol. 18, (1975), pp. 21-31.

[BACL92]Baclace, P.E., "Competitive Agents for Information Filtering", Communications of the ACM, Vol. 35, No. 12, (December 1992), p. 50.

[BACL91]Baclace, P.E., (November 1991). "Personal Information Intake Filtering," Proceedings of Bellcore Workshop on High Performance Information Filtering [Online], p. 3-16, Available FTP: netcom2.netcom.com Directory: pub/peb File: ifilter.ps.

[BAKE62]Baker, F.B., "Information Retrieval Based Upon Latent Class Analysis," Journal of the ACM, Vol. 9, (1962), pp. 512-521.

[BALD94]Baldwin, D. and S.B. Yadav, "The Process of Research Investigation in Artificial Intelligence--A Unified View," forthcoming in IEEE Transaction on Systems, Man, and Cybernetics, 1994, 51 pages.

[BATT89]Batty, D., "Thesaurus Construction and Maintenance: A Survival Kit," Database, Vol. 12, No. 1, (February 1989), pp. 13-20.

[BEID90]Beidleman, C.R., D. Fletcher, and D. Veshosky, "On Allocating Risk: The Essence of Project Finance," Sloan Management Review, Spring 1990, pp. 47-55.

[BELK90]Belkin, N.J., "The Cognitive Viewpoint in Information Science," Journal of Information Science, Vol. 16, No. 1, (1990), pp. 11-15.

[BELK91]Belkin, N.J., "A Methodology for Taking Account of User Tasks Goals and Behavior for Design of Computerized Library Catalogs," SIGCHI Bulletin, Vol. 23, No. 1 (January 1991) pp. 61-65.

[BELK92]Belkin, N.J. And W.B. Croft, "Information Filtering and Information Retrieval: Two Sides of the Same Coin?", Communications of the ACM, Vol. 35, No. 12, (December 1992), pp. 29-38.

[BELK93]Belkin, N.J., P.G. Marchetti, and C. Cool, "BRAQUE: Design of an Interface to Support User Interaction in Information Retrieval," Information Processing and Management, Vol. 29, No. 3, (1993), pp. 325-344.

[BENH93]Benham, P.O., "Developing Organizational Talent: The Key to Performance and Productivity," SAM Advanced Management Journal, Winter 1993, pp. 34-39.

[BOWE92]Bowen, T.F., G. Gopal, G. Herman, T. Hickey, K.C. Lee, W.H. Mansfield, J. Raitz, and A. Weinrib, "The Datacycle Architecture," Communications of the ACM, Vol. 35, No. 12, (December 1992), pp. 71-79.

[BRAD93]Brady, J., "An Examination of Concept Lattices, Types, and Functions in *Roget's International Thesaurus*," Behavior Research Methods, Instruments, and Computers, Vol. 25, No. 2, (1993), pp. 328-332.

[BRAJ87]Brajnik, G., G. Guida, and C. Tasso, "User Modeling in Intelligent Information Retrieval," Information Processing and Management, Vol. 23, No. 4, (1987), pp. 305-320.

[BROO85]Brooks, H.M., P.J. Daniels, and N.J. Belkin, "Problem Descriptions and User Models: Developing an Intelligent Interface for Document Retrieval Systems," Proceedings of Informatics 8: Advances in Intelligent Retrieval, Aslib, London, 1985, pp. 191-214.

[CARL90]Carlson, D.A. and S. Ram, "Hyperintelligence: The Next Frontier," Communications of the ACM, Vol. 33, No. 3, (March 1990), pp. 311-321.

[CHEN94]Chen, P.S., "On Inference Rules of Logic-Based Information Retrieval Systems," Information Processing and Management, Vol. 30, No. 1, (1994), pp. 43-59.

[COHE88]Cohen P. and A. Howe, "How Evaluation Guides AI Research," AI Magazine, Winter 1988, pp. 35-43.

[COHE89]Cohen P. and A. Howe, "Toward AI Research Methodology: Three Case Studies in Evaluation," IEEE SMC, Vol. 19, No. 3, (1989), pp. 634-645.

[COOK79]Cook, T. and D. Campbell, Quasi Experiment: Design and Analysis Issues for Field Settings, Chapter 2: Validity, pp. 37-94, Houghton-Mifflin, Boston, Massachusetts, 1979.

[CRAI43]Craik, K. The Nature of Explanation, Cambridge University Press, Cambridge, London, 1943.

[CROF93]Croft, W. B. and H.R. Turtle, "Retrieval Strategies for HyperText," Information Processing and Management, Vol. 29, No. 3, (1993), pp. 313-324.

[DANI86]Daniels, P.J., "Cognitive Models in Information Retrieval--An Evaluative Review," Journal of Documentation, Vol. 42, No. 4, (December 1986), pp. 272-304.

[DAY92]Day, A.C., "Computer Help in Thesaurus Construction," Literary and Linguistic Computing, Vol. 7, No. 4, (1992), pp. 226-230.

[DEER90]Deerwester, S., S.T. Dumais, G.W. Furnas, T.K. Landauer, R. Harshman, "Indexing by Latent Semantic Analysis," Journal of the American Society for Information Science, Vol. 41, No. 6, (September 1990), pp. 391-407.

[DEER88]Deerwester, S., S.T. Dumais, T.K. Landauer, G.W. Furnas, and L. Beck, "Improving Information Retrieval With Latent Semantic Indexing," Proceedings of the 1988 Annual Meeting of the American Society for Information Science, 1988, pp. 36-40.

[DEJO82]DeJong, G., "An Overview of the FRUMP System," in Strategies for Natural Language Processing, eds. Lehnert, W.G. and M.H. Ringle, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1982, pp. 149-176.

[DEJO79]DeJong, G., "Prediction and Substantiation: A New Approach to Natural Language Processing," Cognitive Science, Vol. 3, (1979), pp. 251-273.

[DENN82]Denning, P.J., "Electronic Junk," Communications of the ACM, Vol. 25, No. 3, (March 1982), pp. 163-165.

[DIET93]Dieterich, H., U. Malinowski, T. Kühme, M. Schneider-Hufschmidt, "State of the Art in Adaptive User Interfaces," in Adaptive User Interfaces, eds. M. Schhneider-Hufshmidt, T. Kühme, and U. Malinowski, Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1993, pp. 13-48.

[DIFF83]Diffenbach, J., "Corporate Environmental Analysis in Large U.S. Corporations," Long Range Planning Journal, Vol. 16, (1983), pp. 107-116.

[DINS90]Dinsmore, P.C., "Ideas, Guidelines, and Techniques for Applying Project Management Solutions in the General Business Arena: Lessons for Executives," Project Management, Vol. 8, No. 1, (February 1990), pp. 33-38.

[DUMA91]Dumais, S.T., "Improving the Retrieval of Information from External Sources," Behavioral Research Methods, Instruments and Computers, Vol. 23, No. 2, (1991), pp. 229-236.

[DUMA88] Dumais, S.T., G.W. Furnas, T.K. Landauer, S. Deerwester, and R. Harshman, "Using Latent Semantic Analysis to Improve Access to Textual Information," CHI '88 Conference Proceedings: Human Factors in Computing Systems, May 1988, pp. 281-285.

[EAST88]Eastman, C.M., "Overlaps in Postings to Thesaurus Terms: A Preliminary Study," Proceedings of the 1988 Annual Meeting of the American Society for Information Science, 1988, pp. 181-184.

[EKME92]Ekmekcioglu, F.C., A.M. Robertson, and P. Willett., "Effectiveness of Query Expansion in Ranked-Output Document Retrieval Systems," Journal of Information Science, Vol. 18, No. 2, (1992), pp. 139-147.

[ELOF91]Elofson, G. and B. Konsynski, "Delegation Technologies: Environmental Scanning with Intelligent Agents," Journal of Management Information Systems, Vol. 8, No. 1, (Summer 1991), pp. 37-62.

[FAHE81]Fahey, L., W.R. King, and V.K. Narayanan, "Environmental Scanning and Forecasting in Strategic Planning: The State of the Art," Long Range Planning Journal, Vol. 14, No. 1, (April 1981), pp. 32-39.

[FARR93]Farrow, Rik, "Satellite Broadcast News," UnixWorld, Vol. 10, No. 12, ((December 1993), pp. 98-99.

[FINE74]Fine, S., "Functional Job Analysis: An Approach to a Technology for Manpower Planning," Personnel Journal, November 1974, pp. 813-818.

[FISC91]Fischer, G. and C. Stevens, "Information Access in Complex, Poorly Structured Information Spaces," Human Factors in Computing Systems: Reaching Through Technology (CHI'91 Conference Proceedings), April 1991, New Orleans, Louisiana, pp. 63-70.

[FOLT90]Foltz, P.W., "Using Latent Semantic Indexing for Information Filtering," ACM/SIGOIS Bulletin, Vol. 11, No. 2/3, (1990), pp. 40-47.

[FOLT92]Foltz, P.W. and S.T. Dumais, "Personalized Information Delivery: An Analysis of Information-Filtering Methods," Communications of the ACM, Vol. 35, No. 12, (December 1992), pp. 51-60.

[FRIS89]Frisse, M.E. and S.B. Cousins, "Information Retrieval from Hypertext: Update on the Dynamic Medical Handbook Project," Hypertext '89 Proceedings, ACM SIGCHI Bulletin, November 1989, pp. 199-212.

[FURN85]Furnas, G.W., "Experience With an Adaptive Indexing Scheme," CHI '85 Proceedings: Human Factors in Computer Systems, April 1985, pp. 131-135.

[FURN88]Furnas, G.W., S. Deerwester, S.T. Dumais, T.K. Landauer, R.A. Harshman, L.A. Streeter, and K.E. Lochbaum, "Information Retrieval Using a Singular Value Decomposition Model of Latent Semantic Structure," Proceeding of the 11th International Conference on Research and Development in Information Retrieval, (June 1988), Grenoble, France, ACM/SIGIR, New York, 1988, pp. 465-480.

[FURN83]Furnas, G.W., T.K. Landauer, L.M. Gomez, and S.T. Dumais, "Statistical Semantics: Analysis of the Potential Performance of Keyword Information Systems," Bell System Technical Journal, Vol. 62, No. 6, (1983), pp. 1753-1806.

[FURN87]Furnas, G.W., T.K. Landauer, L.M. Gomez, and S.T. Dumais, "The Vocabulary Problem in Human-System Communication," Communication of the ACM, Vol. 30, No. 11, (November 1987), pp.964-971.

[GIAN93]Giannotti, A.C. and D.J. Fisher, "Project Information Management System--Another Approach," Journal of Management in Engineering, Vol. 9, No. 1, (January 1993), pp. 52-63.

[GOLD92]Goldberg, D., D. Nichols, B.M. Oki, and D. Terry, "Using Collaborative Filtering to Weave an Information Tapestry," Communications of the ACM, Vol. 35, No. 12, (December 1992), pp. 61-70.

[GRAN89]Grant, P.C., Multiple Use Job Descriptions, Quorum Books, New York, 1989.

[GROS90]Gross, D. and K.J. Miller, "Adjectives in WordNet," International Journal of Lexicography, Vol. 3, No. 4, (Winter 1990), pp. 265-277.

[HALL93]Hall, J. and C.W. Hofer, "Venture Capitalists' Decision Criteria in New Venture Evaluation," Journal of Business Venturing, Vol. 8, No. 1, (January 1993), pp. 25-42.

[HAMB82]Hambrick, D.C., "Environmental Scanning and Organizational Strategy," Strategic Management Journal, Vol. 3, (1982), pp. 159-174.

[HAMB81]Hambrick, D.C., "Specialization of Environmental Scanning Activities Among Upper Level Executives," Journal of Management Studies, Vol. 18, (1981), pp. 299-320.

[HANC92a]Hancock-Beaulieu, M., "Query Expansion: Advances in Research in Online Catalogs," Journal of Information Science, Vol. 18, No. 2, (1992), pp. 99-103.

[HANC92b]Hancock-Beaulieu, M. and S. Walker, "An Evaluation of Automatic Query Expansion in an Online Library Catalogue," Journal of Documentation, Vol. 48, No. 4, (December 1992), pp. 406-421.

[HANS71]Hansen, W., User Engineering Principles for Interactive Systems, AFIPS Press, Montvale, New Jersey, 1971.

[HOWE92]Howes, R., W. Little, and D. Fong, "Dynamic Project Control Utilizing a New Approach to Project Control," Building Research and Information, Vol. 20, No. 5, (September 1992), pp. 291-294.

[HREB85]Hrebiniak, L.L. and W.F. Joyce, "Organizational Adaptation: Strategic Choice and Environmental Determinism," Administrative Science Quarterly, Vol. 30, (1985), pp. 336-349.

[INGR92]Ingram, P., "Career Mapping -- An Edge on Career Advancement," Recruiter Journal, November 1992, pp. 12-15.

[JACO90]Jacobs, P.S. and L.F. Rau, "SCISOR: Extracting Information from On-line News," Communications of the ACM, Vol. 33, No. 11, (November 1990), pp. 88-97.

[JENN92]Jennings, D.F. and J.R. Lumpkin, "Insights Between Environmental Scanning Activities and Porter's Generic Strategies: An Empirical Analysis," Journal of Management, Vol. 18, No. 4, (1992), pp. 791-803.

[JENN94]Jennings, R., Access 2 Developer's Guide, Sams Publishing, Indianapolis, Indiana, 1994.

[JOHN80]Johnson-Laird, P.N., "Mental Models in Cognitive Science," Cognitive Science, Vol. 4, (1980), pp. 71-115.

[JOHN83]Johnson-Laird, P.N., Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness, Cambridge University Press, Cambridge, London, 1983.

[JONE93]Jones, S., "A Thesaurus Data Model for an Intelligent Retrieval System," Journal of Information Science, Vol. 19, No. 3, (1993), pp. 167-178.

[KASS89]Kass, R. and T. Finin, "The Role of User Models in Cooperative Interactive Systems," International Journal of Intelligent Systems, Vol. 4, (1989), pp. 81-112.

[KERN92]Kern, A.G., "Clearing the Hurdles to Project Success," InTech, Vol. 39, No. 7, (July 1992), p. 17.

[KHAZ91]Khazanchi, D., "Subjective Understanding of Ill-Structured Problems: An Information Systems Perspective," Ph.D. Dissertation, Texas Tech University, 1991.

[KING87]King, W.R., "Decision Support Systems," in Strategic Planning and Management Handbook, eds. W.R. King and D.I. Cleland, Van Nostrand Reinhold Company, New York, 1987, pp. 422-435.

[KRIS93]Kristensen, J., "Expanding End-User's Query Statements for Free Text Searching with a Search-Aid Thesaurus," Information Processing and Management, Vol. 29, No. 6, (1993), pp. 733-744.

[LAUF90]Laufer, Alexander, "Essentials of Project Planning: Owner's Perspective," Journal of Management in Engineering, Vol. 6, No. 2, (April 1990), pp. 162-176.

[LEE94]Lee, J.H., M.H. Kim, and Y.J. Lee, "Ranking Documents in Thesaurus-Based Boolean Retrieval Systems," Information Processing and Management, Vol. 30, No. 1, (1994), pp. 79-91.

[LEVE84]Levesque, H.J., "Foundations of a Functional Approach to Knowledge Representation," Artificial Intelligence, Vol. 23, (1984), pp. 155-212.

[LEVI83]Levine, E.L., R.A. Ash, H. Hall, and F. Sistrunk, "Evaluation of Job Analysis Methods by Experienced Job Analysts," Academy of Management Journal, Vol. 28, No. 1 (1983), pp. 339-348.

[LOEB92a]Loeb, S., "Architecturing Personalized Delivery of Multimedia Information," Communications of the ACM, Vol. 35, No. 12, (December 1992), pp. 39-48.

[LOEB92b]Loeb, S., R. Hill, and T. Brinck, "Lessons from *LyricTime*: A Prototype Multimedia System," Proceedings of the IEEE Workshop on Multimedia '92, (Monterey, California, April) IEEE, New York, 1992, pp. 106-113.

[LUHN58]Luhn, H.P., "A Business Intelligence System," IBM Journal of Research and Development, Vol. 2, 1958, pp. 314-319.

[MAGJ93]Magjuka, R.J., "The 10 Dimensions of Employee Involvement," Training and Development, April 1993, pp. 61-67.

[MALO87]Malone, T.W., K.R. Grant, F.A. Turbank, S.A. Brobsy, and M.D. Cohen, "Intelligent Information-Sharing Systems," Communications of the ACM, Vol. 30, No. 5, (May 1987), pp. 390-402.

[MANE88]Manese, W.R., Occupational Job Evaluation, Quorum Books, New York, 1988.

[MAUL89]Mauldin, M.L., "Information Retrieval by Text Skimming," Ph.D. Dissertation, Carnegie Mellon University, 1989.

[MAUL91]Mauldin, M.L., (1991), "Retrieval Performance in FERRET," [Online]. Available FTP: nl.cs.cmu.edu Directory:/usr/mlm/ftp/pubs File:sigir91.ps.Z.

[MAZE91]Maze, M. and D. Mayall, The Enhanced Guide for Occupational Exploration, JIST Works, Inc, Indianapolis, Indiana, 1991.

[MCCA92]McCann, J.E. and L. Gomez-Mejia, "Going 'On-Line' in the Environmental Scanning Process," IEEE Transactions on Engineering Management, Vol. 39, No. 4, (November 1992), pp. 394-399.

[MCKI90]McKim, R.A., "Project Control--Back to Basics," Cost Engineering, Vol. 32, No. 12, (December 1990), pp. 7 -11.

[MCMA89]McMath, C.F., R.S. Tamaru, and R. Rada, ""A Graphical Thesaurus-Based Information Retrieval System," International Journal of Man-Machine Studies, Vol. 31, No. 2, (August 1989), pp. 121-147.

[MESC91]Mesch, A.H., "Developing an Effective Environmental Assessment Function," Managerial Planning, Vol. 32, No. 5, (March/April 1984), pp. 17-22, reprinted in Strategic Management Practice, eds. J.A. Pearce II and R.B. Robinson, Jr., Richard B. Irwin, Inc., Boston, Massachusetts, 1991.

[MILL89]Miller, D. "Matching Strategies and Strategy Making: Process, Content, and Performance," Human Relations, Vol. 42, (1989), pp. 241-260.

[MILL90a]Miller, G.A., "Nouns in WordNet: A Lexical Inheritance System," International Journal of Lexicography, Vol. 3, No. 4, (Winter 1990), pp. 245-264.

[MILL90b]Miller, G.A., R. Beckwith, C. Fellbaum, D. Gross, and K.J. Miller, "Introduction to WordNet: An On-line Lexical Database," International Journal of Lexicography, Vol. 3, No. 4, (Winter 1990), pp. 235-244.

[MINT76]Mintzberg, H., "Planning on the Left Side and Managing On the Right," Harvard Business Review, Vol. 54, (July-August, 1976), pp. 49-58.

[MITR85]Mitroff, I.I., "Two Fables for Those Who Believe in Rationality," Technological Forecasting and Social Change, Vol. 28, (1985), pp. 195-202.

[MORR88]Morris, A.H., "Supporting Environmental Scanning and Organizational Communication with the Processing of Text: The Use of Computer-Generated Abstracts," Ph.D. Dissertation, Texas Tech University, 1988.

[MYAE90]Myaeng, S.H. and R.R. Korfhage, "Integration of User Profiles: Models and Experiments in Information Retrieval," Information Processing and Management, Vol. 26, No. 6, (1990), pp. 719-738.

[NARA87]Narayanan, V.K. and L. Fahey, "Environment Analysis for Strategy Formulation" in Strategic Planning and Management Handbook, eds. W.R. King and D.I. Cleland, Van Nostrand Reinhold Company, New York, 1987, pp. 147-175.

[NEWE82]Newell, A., "The Knowledge Level," Artificial Intelligence, Vol. 18, (1982), pp. 82-127.

[NEWE93]Newell, A., "Reflections on the Knowledge Level," Artificial Intelligence, Vol. 59, (1993), pp. 31-38.

[NEWE90]Newell, A., Unified Theories of Cognition, Harvard University Press, Cambridge, Massachusetts, 1990.

[NEWS93]NewsEDGE at a Glance. Waltham, Massachusetts: Desktop Data, Inc., 1993.

[NORM86]Norman, D.A., "Cognitive Engineering," in User Centered System Design, eds. Norman, D.A. and S.W. Draper, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1986, pp. 31-62.

[NOTE93]Notess, G.R., "Usenet: Taming the E-mail Flood," Online, Vol. 17, No. 4, (July 1993), pp. 86-88.

[OLEA88]O'Leary, D., "Software Engineering and Research Issues in Accounting Information Systems," Journal of Information Systems, (Spring 1988), pp. 24-38.

[PALM88]Palmquist, R.A. and B. Balakrishnan, "Using a Continuous Word Association Test to Enhance a User's Description of an Information Need: A Quasi-Experimental Study," Proceedings of the 1988 Annual Meeting of the American Society for Information Science, 1988, pp. 160-163.

[PINT90a]Pinto, J.K., "Project Implementation Profile: A Tool to Aid Project Tracking and Control," International Journal of Project Management, Vol. 8, No. 3, (August 1990), pp. 173-182.

[PINT90b]Pinto, J.K. and J.E. Prescott, "Planning and Tactical Factors in the Project Implementation Process," Journal of Management Studies, Vol. 27, No. 3, (May 1990), pp. 305-327.

[POLL93]Pollard, R., "A Hypertext-Based Thesaurus as a Subject Browsing Aid for Bibliographic Databases," Information Processing and Management, Vol. 29, No. 3, (1993), pp. 345-357.

[POPP61]Popper, K.R., The Logic of Scientific Discovery, Science Editions, New York, 1961.

[RAHB91]Rahbar, F.F., J.K. Yates, and G.R. Spencer, "Project Management Knowledge Engineering System," Cost Engineering, Vol. 33, No. 7, (July 1991), pp. 15-24.

[RAIN93]Rainbird, H. and M. Maguire, "When Corporate Need Supersedes Employee Development," Personnel Management, February 1993, pp. 34-38.

[RAM92]Ram, A., "Natural Language Understanding for Information-Filtering Systems," Communications of the ACM, Vol. 35, No. 12, (December 1992), pp. 80-81.

[RHYN87]Rhyne, L.C., "Information System Support," in Strategic Planning and Management Handbook, eds. W.R. King and D.I. Cleland., Van Nostrand Reinhold Company, New York, 1987, pp. 387-405.

[RICH79]Rich, E., "User Modeling via Stereotypes," Cognitive Science, Vol. 3, (1979), pp. 329-354.

[RICH83]Rich, E., "Users Are Individuals: Individualizing User Models," International Journal of Man-Machine Studies, Vol. 18, No. 3, (1983), pp. 199-214.

[RICH91]Rich, E. and K. Knight, Artificial Intelligence, McGraw-Hill, Inc., New York, 1991.

[ROBE88]Robert, M., The Strategist CEO, Quorum Books, New York, 1988.

[ROBE90]Robertson, S.E., "On Term Selection for Query Expansion," Journal of Documentation, Vol. 46, No. 4, (December 1990), pp. 359-364.

[SALT90]Salton, G. and C. Buckley, "Improving Retrieval Performance by Relevance Feedback," Journal of the American Society for Information Science, Vol. 41, No. 4, (1990), pp. 288-297.

[SAUN90]Saunders, R.G., "Project Management in R&D: the Art of Estimating Development Project Activities," Project Management, Vol. 8, No. 1, (February 1990), pp. 45-50.

[SCHU87]Schultz, R.L., D.P. Slevin, and J.K. Pinto, "Strategy and Tactics in a Process Model of Project Implementation," Interfaces, Vol. 17, No. 3, (1987), pp. 34-46.

[SCHW88]Schwenk, C.R., "The Cognitive Perspective of Strategic Decision Making," Journal of Management Studies, Vol. 25, No. 1, (January 1988), pp. 41-55.

[SHAN75]Shannon, R.E., Systems Simulation: The Art and Science, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

[SHER93]Sherman, E.H. and E.H. Shortliffe, "A User-Adaptable Interface to Predict Users' Needs," in Adaptive User Interfaces, eds. M. Schhneider-Hufshmidt, T. Kühme, and U. Malinowski, Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1993, pp. 285-315.

[SHET94]Sheth, B.D., "A Learning Approach to Personalized Information Filtering," Master's Thesis, Massachusetts Institute of Technology, 1994.

[SNYD81]Snyder, N., "Environmental Volatility, Scanning Intensity, and Organizational Performance," Journal of Contemporary Business, Vol. 10, (1981), pp. 5-17.

[SOMM89]Sommerville, I, "Chapter 21: Testing Techniques," Software Engineering, Addison-Wesley, Reading, Massachusetts, 1989, pp. 421-436.

[STAD91]Stadnyk, I. and R. Kass, "Modeling Decision Making of Usenet News Readers," Technical Report CFAR-91-003, EDS Center for Advanced Research, 1991.

[STAD92]Stadnyk, I. and R. Kass, "Modeling Users' Interests in Information Filters," Communications of the ACM, Vol. 35, No. 12, (December 1992), pp. 49-50.

[STEV92]Stevens, C., "Automating the Creation of Information Filters," Communications of the ACM, Vol. 35, No. 12, (December 1992), p. 48.

[STOF82]Stoffels, J.D., "Environmental Scanning for Future Success," Managerial Planning, Vol. 31, No. 3, (November-December 1982), pp. 4-12.

[STUB82]Stubbart, C., "Are Environmental Units Effective?," Long Range Planning Journal, Vol. 15, (1982), pp. 139-145.

[TERR77]Terry, P.T., "Mechanisms for Environmental Scanning," Long Range Planning, Vol. 10, (June 1977), pp. 2-9.

[TESK89]Teskey, F.N., "User Models and World Models for Data, Information, and Knowledge," Information Processing and Management, Vol. 25, No. 1 (1989), pp. 7-14.

[WANG85]Wang, Y.C., J. Vandendorpe, and M. Evens, "Relational Thesauri in Information Retrieval," Journal of the American Society for Information Science, Vol. 36, No. 1, (1985), pp. 15-27.

[WOLF88]Wolff, R.H., The Complete Portfolio of Prewritten Job Descriptions, Asher-Gallant Press, Westbury, New York, 1988.

[WOTH92]Wotherspoon, I., "Historical Thesaurus Database Using Ingres," Literary and Linguistic Computing, Vol. 7, No. 4, (1992), pp. 218-225.

[YADA92]Yadav, S.B. and D. Khazanchi, "Subjective Understanding in Strategic Decision Making: An Information Systems Perspective," Decision Support Systems, Vol. 8, (1992) pp. 55-71.

[YAZI93]Yazici, H. and R. Kluczny, "Information Display Modes and User Cognitive Profiles: Interaction Effects on the Decision Making Process," Journal of Computer Information Systems, Summer 1993, pp. 41-54.

APPENDIX: USER'S GUIDE TO THE HOLISTIC PROFILE FILTERING SYSTEM

Introduction

The holistic profile filtering system (HPFS) provides the user with an enhanced filtering system that is capable of gathering data that impacts the user's professional capabilities, project management skills, awareness of industry threats and opportunities, and level of client support. HPFS is an object-oriented application, designed around a database engine. The database application development environment was provided by Microsoft Access 2.0.

Hardware/Software Requirements

Because the foundation of this software package is Access 2.0, the system requirements are extremely demanding. HPFS is a resource-intensive application, and requires an 80486-based or better IBM-compatible microcomputer with a fixed disk drive, capable of running Windows 3.0[3] or higher in standard or enhanced mode. HPFS also requires a system with at least 8 MB of RAM. Although it will operate on a machine with only 4 MB of RAM, system performance will be severely degraded. The amount of RAM has a greater impact on system performance than does the processor speed or power.

In the current release, HPFS will only run under Access 2.0, and is not a stand-alone application. A complete installation of Access 2.0 requires approximately 20 MB of free disk space, and the HPFS requires another 2 to 3 MB. A permanent Windows swap file of at least 9 MB is required for best performance.

A mouse or trackball is required in the current release of HPFS. User screens incorporate colors, and therefore a high-resolution color monitor is recommended.

Initiating and Exiting the HPFS System

In order to initiate the HPFS, Microsoft Access must be running. To start Microsoft Access and open the HPFS, perform the following steps:

---

[3]Windows is a trademark of Microsoft Corporation.

1. Double-click the Access icon in the Access application group.
2. Double-click the Open Database button of the toolbar (a small folder), or choose **F**ile from the main menu bar, followed by **O**pen Database from the **F**ile menu.
3. From the **O**pen Database dialog, double-click HPFS in the Directories list to change to the HPFS directory.
4. From the subsequent window double-click on hpfs.mdb in the File Name list to open the HPFS database.

This initiates the HPFS application, which will display an introductory screen and wait for the user to click the Begin button. When the Begin button is clicked the main menu, which is shown in Figure A.1, is displayed. To exit the HPFS double-click the End Program button. The system will exit the HPFS and then provide the option of exiting Access as well, as seen in Figure A.2.

The HPFS User Interface

The HPFS interface is relatively straightforward. Every option available to the user is associated with a button. If the label or icon associated with the button is unclear to the user, clicking once on the button will cause a brief description of the button's function to appear in the status bar at the bottom of the screen. In order to initiate the action associated with the button the user must double-click on the button.

Using HPFS Help

The HPFS is equipped with a full-featured help system based on the Windows 3.1 WinHelp Engine. Help can be accessed in a variety of ways. The most familiar technique for accessing the help system is by hitting the F1 key. In addition, many of the forms are also equipped with a button that has a question mark icon. Both methods initiate the context-sensitive help system.

Context-sensitive help tries to anticipate the user's need for information by displaying help windows related to either the current screen or the function that the user is attempting to perform. For example, if the user double-clicks on the

Generate Profile button and the Overview Screen is displayed, the use can hit F1 or double-click the help button to get information about the purpose of the Overview Screen. If the user is unsure about the function of any of the buttons on the screen, even after reading the description in the status bar, clicking the button and hitting F1 will display a help window describing the button's purpose.

Most of the help windows include hot spots that offer additional information about the topic. Hot spots with dotted underlines display definition windows that define terms used in the window. Hot spots with solid underlines jump to another help window for that topic. A sample help screen containing a hot spot appears in Figure A.3.

Additional information on using the help feature can be obtained by clicking on the menu item Help when any help window is displayed, and selecting the option How to Use Help. Information about the table of contents, searching, and help history can be obtained from this option.

HPFS Input Screens

The input screens associated with profile generation or maintenance consist of a series of labeled input fields and a set of control buttons. While information about any of these features is available through help, a brief description is in order.

The input fields can be one of two types: text boxes or combo boxes. Data is entered in a text box by typing each entry, separating multiple entries by commas. Input in a field is terminated by hitting the enter key or by clicking on a different input field. Combo boxes are text boxes with an associated drop-down list. A combo box accepts typed entries, but also provides the option of selecting entries from the drop-down list. Clicking the combo box's arrow will make the drop-down list appear. The user can select an entry from the list by clicking on it. Unlike standard Access combo boxes, an HPFS combo box allows multiple entries, either typed, selected from the list, or both. A sample input screen is shown in Figure A.4.

Using the Program

In order to use the HPFS the user must first establish a user profile. After creating the profile the system can be used to filter data. To do so the user must first import the data and then filter it. After the data is filtered it can be read and evaluated. Maintenance to the user profile and the provided templates and synonyms can be performed at any time.

Creating a Profile

The process of creating a profile is initiated by double-clicking the Generate Profile button, shown in the submenu in Figure A.5. The Overview Screen shown in Figure A.6 appears first, and requires the user to enter a job title, project type, industry type, and client organization. Selections can be made from the drop-down list. The subsequent screens represent each segment of the holistic profile. Careful thought should be given when creating the profile, because it is the comprehensiveness of the profile that determines the effectiveness of the filtering process. Each field on the input form is labeled with a keyword category. The current input field is highlighted, and the status bar at the bottom of the screen provides a more detailed description of the type of keywords expected. Some input fields are text boxes, in which case the user enters each keyword, separating multiple keywords by commas. Other input fields are combo boxes, in which case the user can either type keywords or select them from the drop-down list that appears when the user clicks on the arrow. Combo boxes allow the user to enter multiple keywords, either typed or selected from the list, or both. The drop-down list for each field is made up of a template of potential keywords to assist the user in generating the profile. An example of such a template can be seen in Figure A.4.

Each form is provided with user-defined categories. These are included in case the user wishes to include keywords that do not fit neatly into other categories. The user can replace the category label "User-defined category" with the label of his or her choosing. The next time the form is opened the new label will be included.

The operations that can be performed on the input forms are initiated from the form-control buttons.

Form-Control Buttons

The form-control buttons perform various functions, such as saving the entries, clearing the form, undoing the clear operation, initiating help, quitting the current operation, and moving to the next screen or the previous screen.

Save Button

The Save button is used to save all of the entries on the form. In most cases, once the entries are saved the form is automatically cleared.

If one of the entered keywords has domain-specific synonyms available, then the user is alerted and then is presented with a list of synonyms from which to make a selection. The user can select any or none of the offered synonyms. The user is then given the option of retaining or discarding the original entry. A sample synonym screen is shown in Figure A.7.

Clear Button

The Clear button is used to clear all of the entries on the input form. If a form is cleared by mistake, the Undo button can be used to restore it.

Undo Button

The Undo button is used to restore the screen after a Clear operation has been performed.

Help Button

The Help button invokes the context sensitive help for the current screen.

Quit Button

The Quit button is used to terminate the current operation and return to the most recent menu.

Next Button

The Next button is used to move to the next screen. If entries have been made on the current screen but not saved, the user is given the option of saving them before moving on.

Back Button

The Back button is used to move to the previous screen. If entries have been made on the current screen but not saved, the user is given the option of saving before moving on.

Importing Data

In order to provide the system with data to be filtered, that data must first be imported into the system. This process, which is initiated by the Import Articles button on the Main Menu shown in Figure A.1, converts standard text files into a form that is usable by the HPFS. When the Import Articles button is double-clicked, the user is asked for the date of the articles to be imported, as in Figure A.8. The user can accept the default, which is the current date, or enter a different date. The process of importing and converting the articles then takes place, and is followed by a message that indicates the number of articles that were imported.

Filtering the Articles

In order to initiate the information filtering process the user double-clicks the Filter News button on the Main Menu shown in Figure A.1. This function searches each article for an occurrence of any of the keywords that appear in the user profile. If a keyword is found then the article is retained. Otherwise it is discarded. When the Filter News button is double-clicked, the user is asked for the date of the articles to be filtered, as shown in Figure A.9. The user can accept the default, which is the current date, or enter a different date. The process of filtering

the articles then takes place, and is followed by a message that indicates the number of articles that were retained.

Reading the News

In order to read the articles that were retained during the filtering process, the user double-clicks on the Read News button shown on the menu in Figure A.5. This causes the system to cycle through the set of retained articles, allowing the user to read and evaluate each in turn.

When an article is displayed, every line that contains one of the keywords from the user profile is highlighted in yellow. This allows the user to quickly scan the article to determine why it was selected and if it is of interest. An example of the article display screen is shown in Figure A.10. When the user finishes reading the article, one of several functions can be selected from the form-control buttons.

Evaluate Button

The Evaluate button closes the viewing screen and enters the article evaluation screen.

Delete Button

The Delete button allows the user to flag an article for deletion. When the user evaluates an article, an option appears that allows the user to save or delete the article. If the article is saved and the user later wants to delete it, the Delete button can be used. A message stating the current state of the delete flag appears.

Undelete Button

The Undelete button is used to turn off the delete flag in the event that it was turned on in error. A message stating the current state of the delete flag appears.

Help Button

The Help button invokes the context sensitive help for the current screen.

Quit Button

The Quit button is used to terminate the current operation and return to the most recent menu.

Evaluating Articles

When the Evaluate button is double-clicked, the evaluation screen shown in Figure A.11 appears. The user is asked if the preceding article was useful, and is also asked to rate the article. To rate the article the user selects one of the options from the drop-down list. The user's rating is very important, because it is that rating that helps the system to determine how well each keyword in the profile is performing. The user should be careful to provide an accurate rating to each article in order to enhance future filtering performance.

Several buttons appear on the evaluation form. While some are familiar by now, others have a slightly different function.

Save Button

The Save button is used to save the user's evaluation.

Help Button

The Help button invokes the context sensitive help for the evaluation screen.

Quit Button

The Quit button is used to terminate the current operation and return to the most recent menu, but only if the article has been evaluated and the evaluation has been saved.

Next Button

The Next button will display the next retained article. If the evaluation has not been saved then the Next operation is aborted.

Back Button

The Back button is used to review the current article. If the user did not read the article closely enough to evaluate it, double-clicking the Back button will re-display the article.

Maintenance

Maintenance capabilities are provided for the user profile, the templates, and the synonym sets. It is unlikely that any but the most advanced user will find it necessary to perform template or synonym maintenance, but the option is available if needed.

The primary types of maintenance that can be performed include creation, modification, addition, and deletion. To create a new profile the user double-clicks the Generate Profile button. To create a new template file the user can double-click the Template File Generation button. To create a new synonym file the user double-clicks the Synonym File Generation button. These options are shown in Figure A.5.

Likewise, the user can modify, add, or delete entries in an existing profile, template, or synonym set. To perform maintenance on the profile the user double-clicks the Profile Maintenance button. To perform maintenance on a template file the user can double-click the Template File Maintenance button. To perform maintenance on a synonym file the user double-clicks the Synonym File Maintenance button. Double-clicking any of these buttons results in a menu similar to that shown in Figure A.12, offering the user the choice of performing modifications, additions, or deletions.

Adapting the Profile

The HPFS provides a feature to assist the user in maintaining the most useful profile possible. The performance of each keyword in the profile is tracked, based

on the user's article evaluations. When the Adapt Profile button is double-clicked, every keyword in the profile is compared to a set of performance thresholds. If the keyword has retrieved only a few documents, or if the keyword consistently retains documents that receive poor ratings, then this is pointed out to the user, who can then decide whether to remove or retain the keyword.

Printing

Various printing features are available. Double-clicking the Print Profile button provides a listing of every keyword in the profile and how it is performing. The Print Input Forms button results in the submenu shown in Figure A.13, which allows the user to select the input form to be printed. Having a copy of the keyword categories might be helpful when the initial profile is being created.

Tips

The effectiveness of the information filtering process is determined by the quality of the profile that guides it. The user can influence this quality in two ways. First, devote a great deal of thought when creating your profile. Let the keyword categories provide guidance, carefully review the template offerings, consider the synonyms that are suggested, and don't be afraid to add new keyword categories. Above all, think carefully when setting up your profile.

The second way in which the user influences the effectiveness of the system is in evaluating the articles. This article evaluation, if done accurately, helps the Adapt Profile feature to alert the user to poor keyword selection. In this way keywords that retain bad articles can be removed and less unwanted material is allowed to pass though the filter.

The Holistic Profile Filtering System is a tool that is intended to screen incoming data and discard information that is of no use to the user. By setting up a comprehensive yet discriminating profile the HPFS can reduce information overload and assist the user in becoming more effective and more efficient.

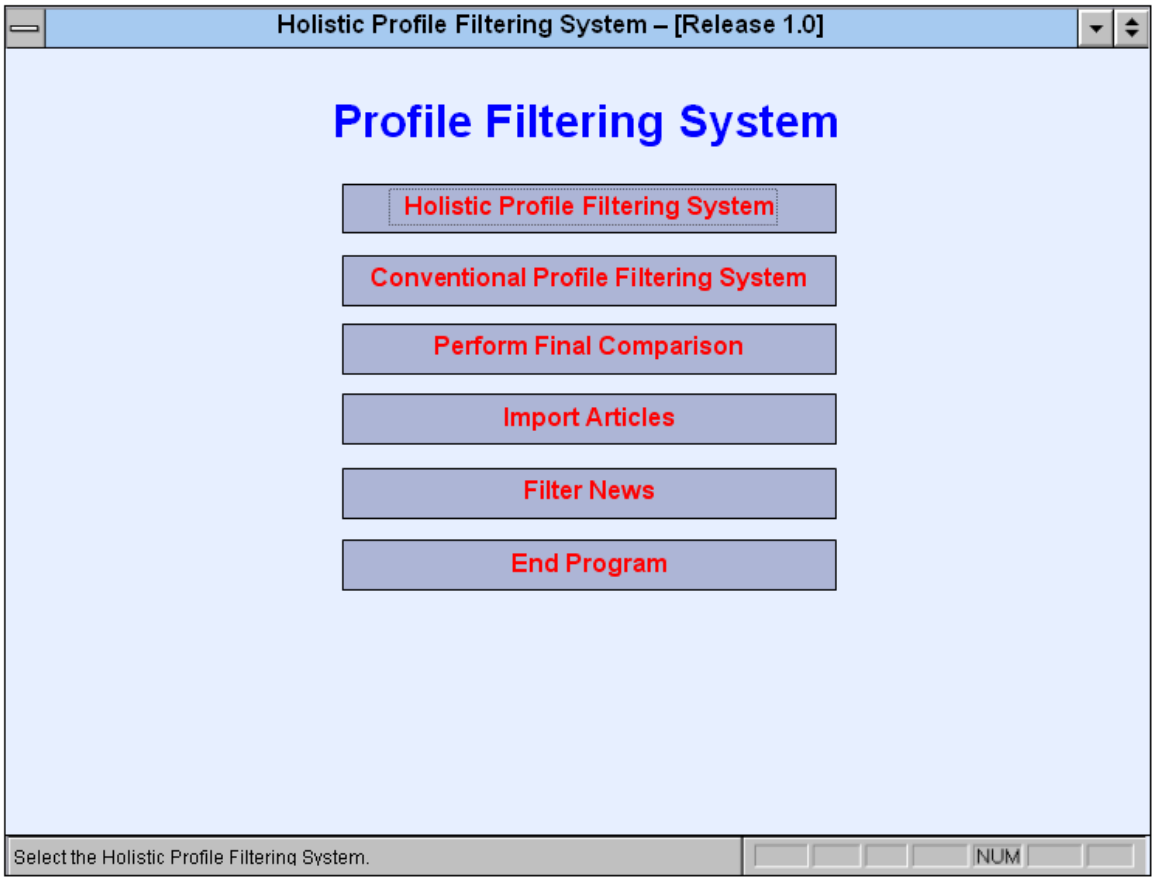

Figure A1. Holistic Profile Filtering System Main Menu.

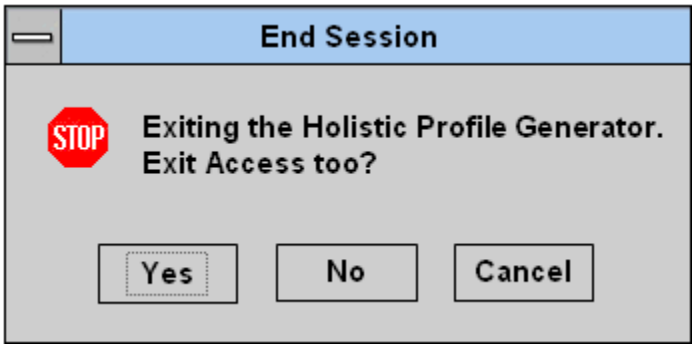

Figure A2. HPFS Exit Option.

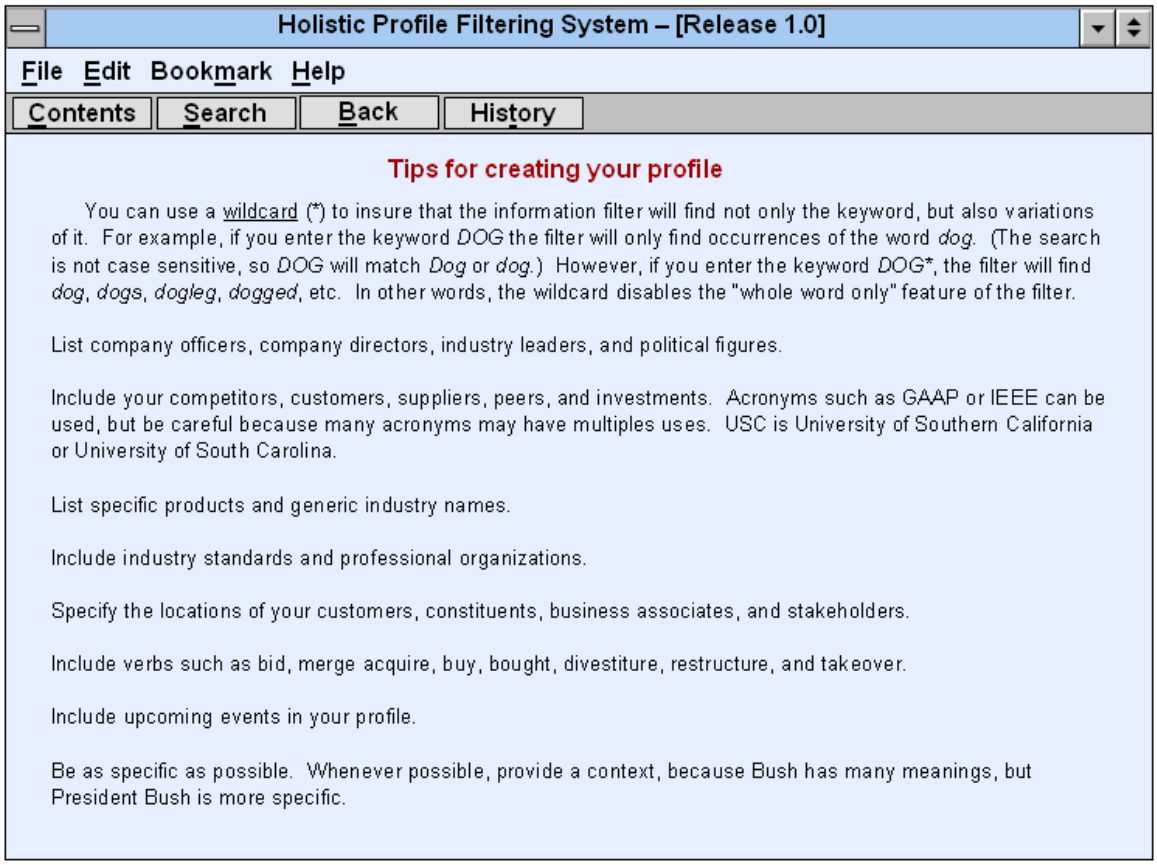

Figure A.3. Sample Help Screen.

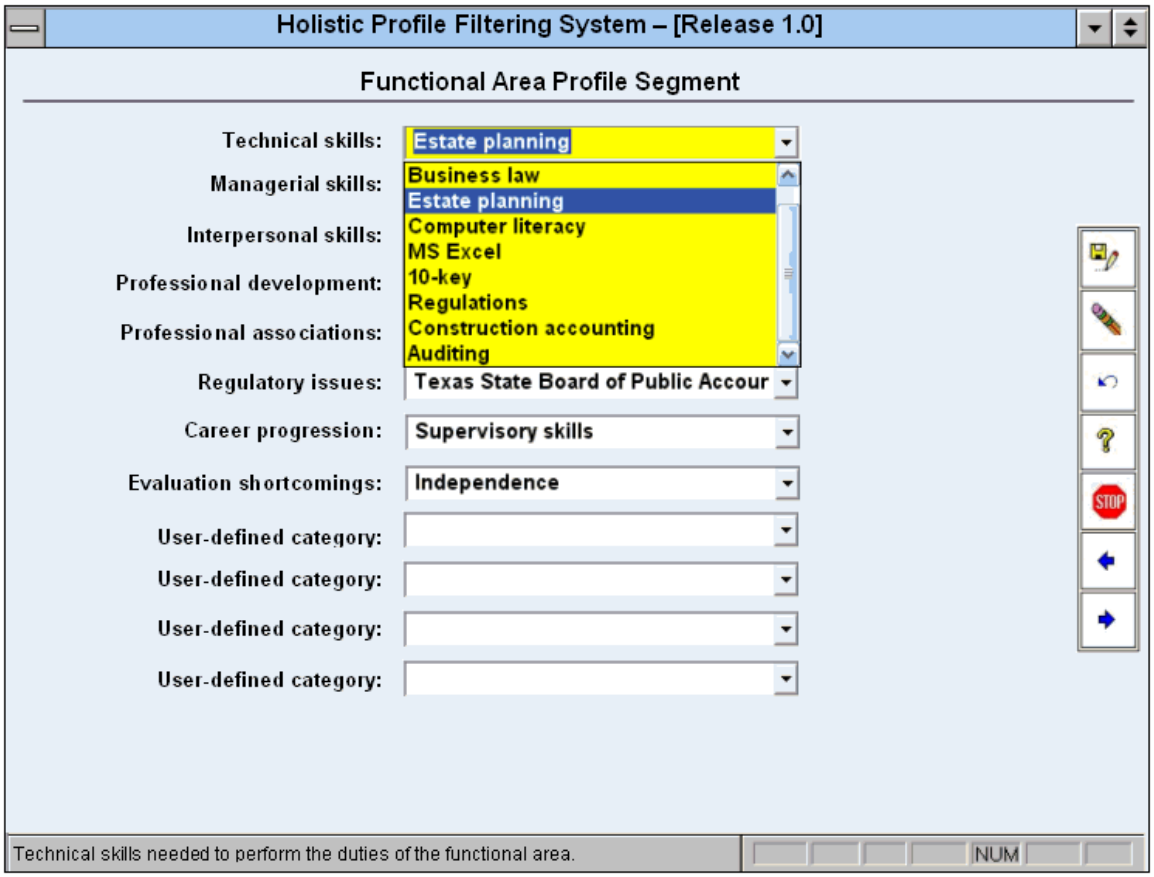

Figure A.4. Input Screen with Templates.

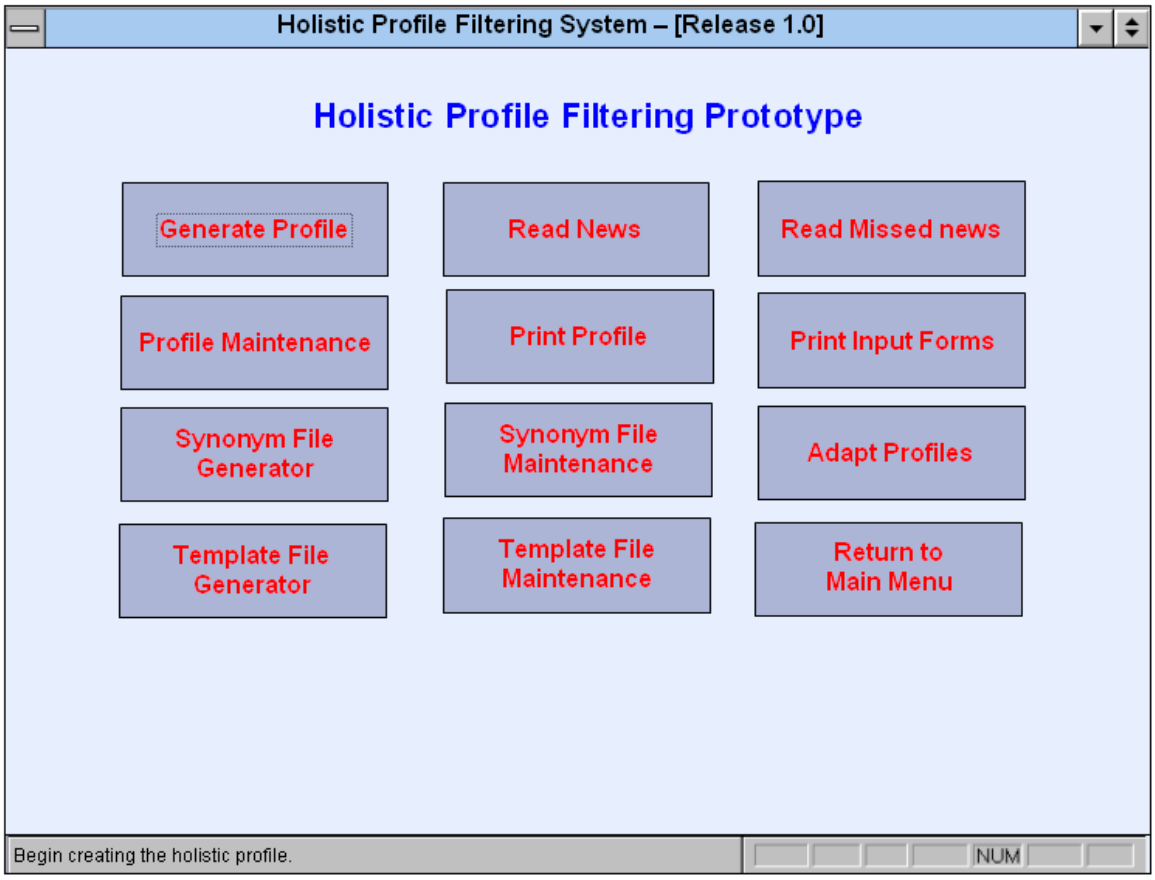

Figure A.5. HPFS Submenu.

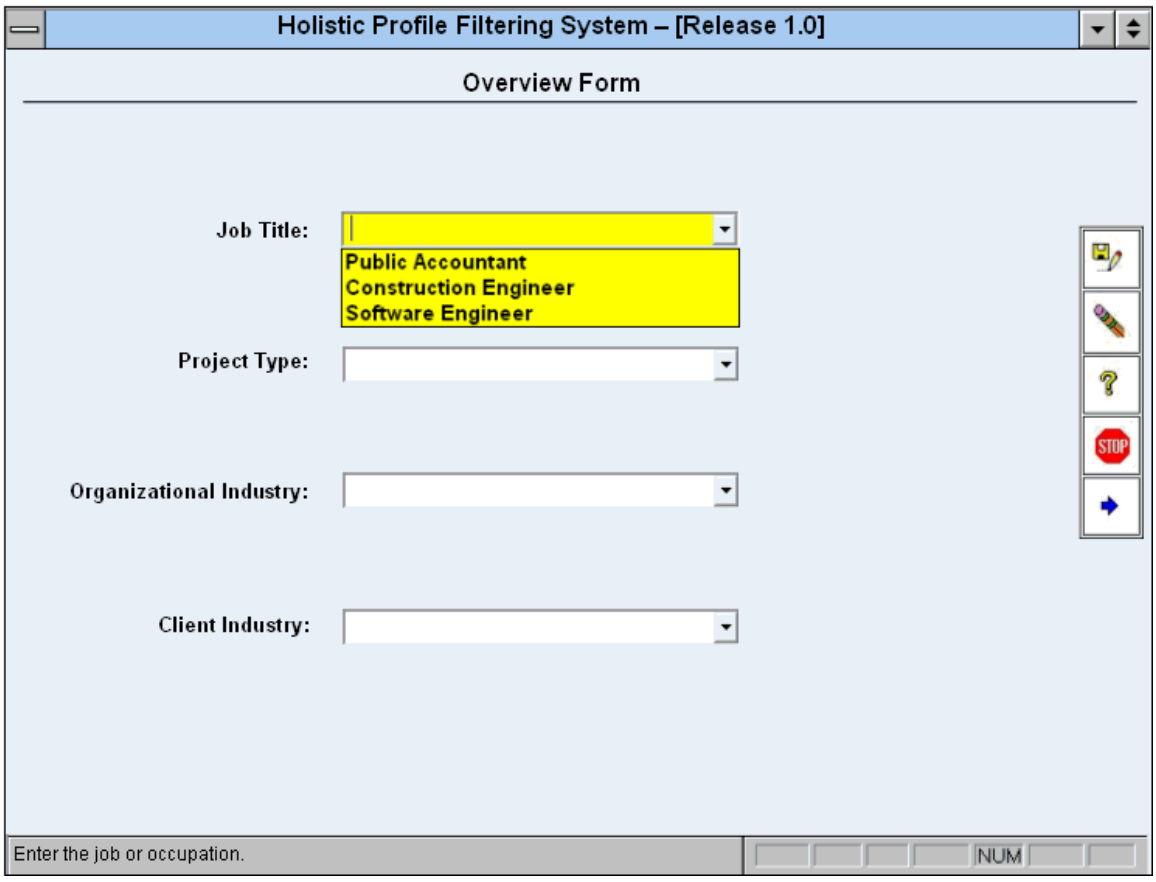

Figure A.6. Overview Form.

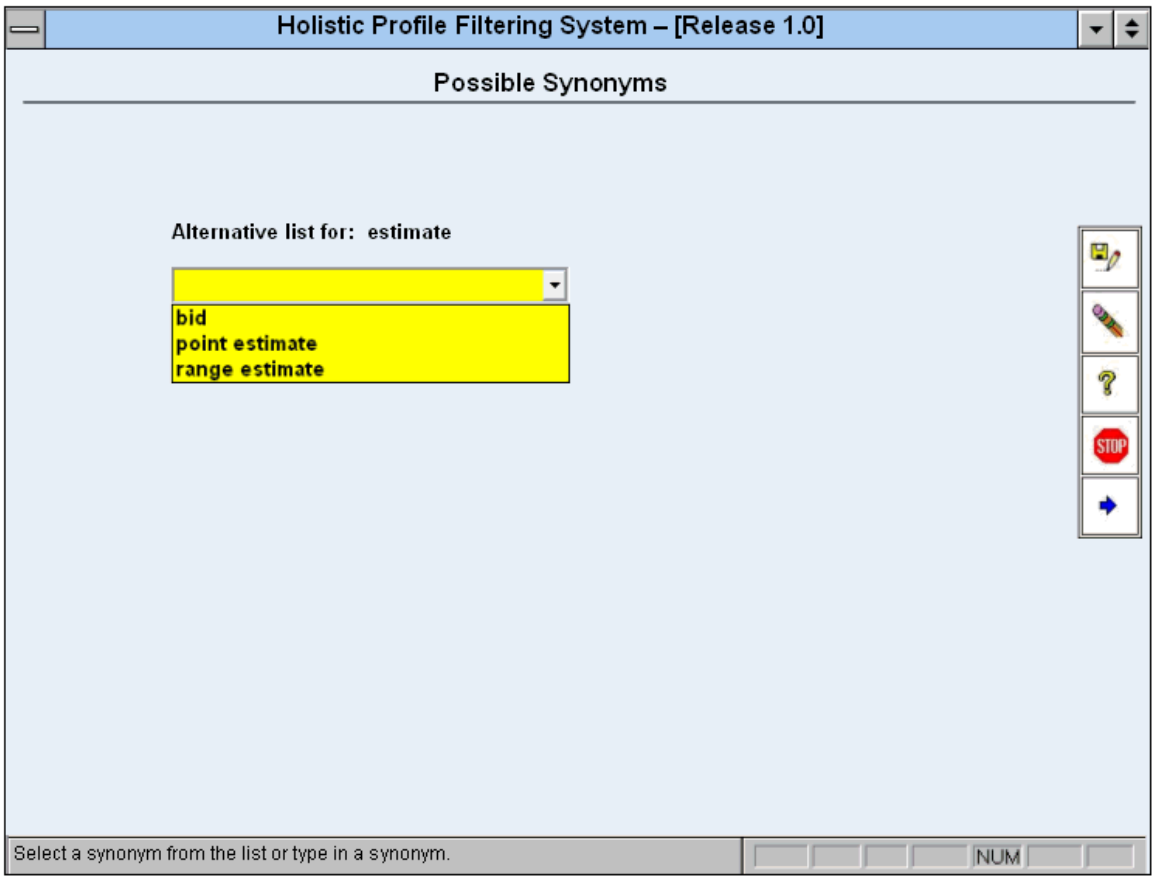

Figure A.7. Synonym Screen.

Figure A.8. Import Articles Screen.

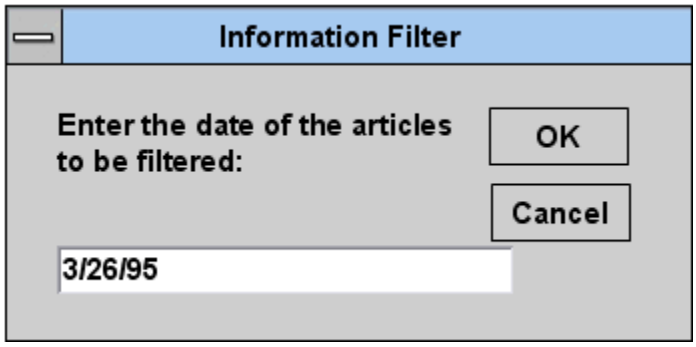

Figure A.9. Filter News Screen.

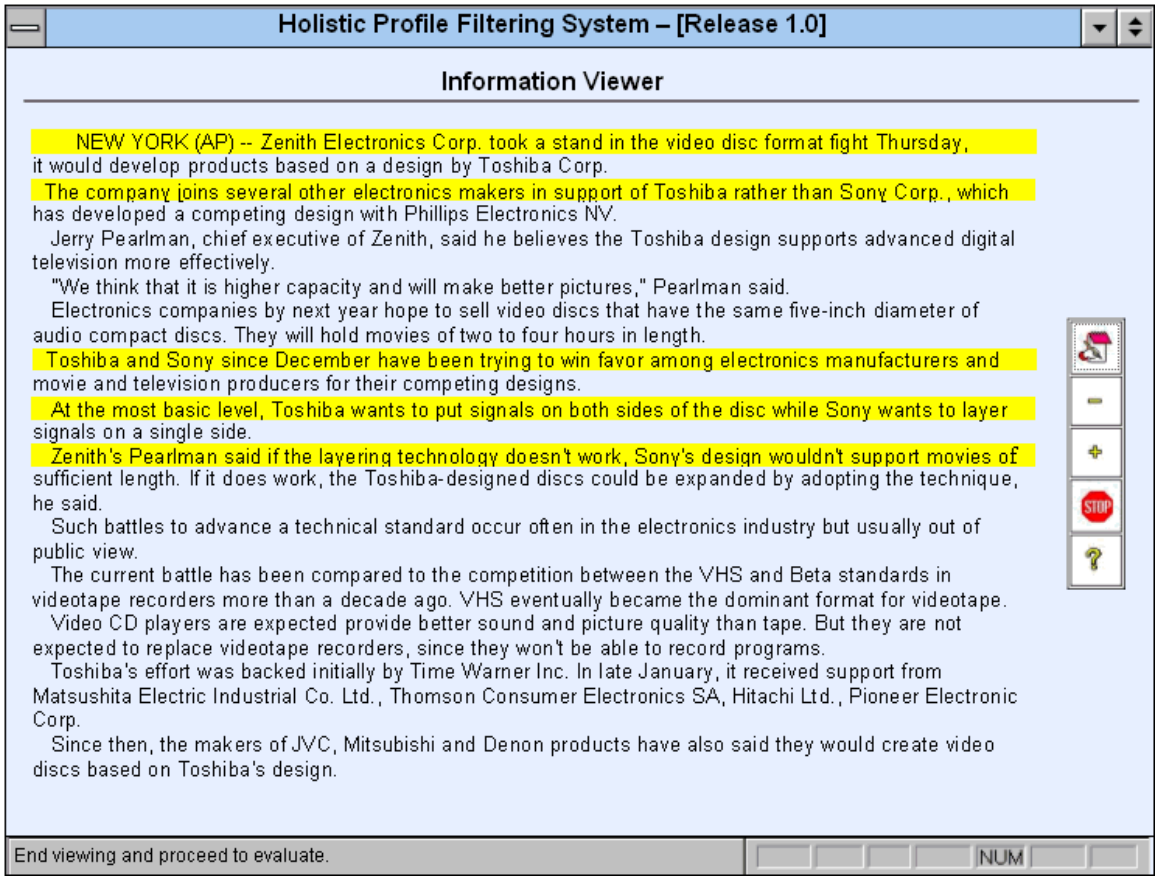

Figure A.10. Sample Article Display.

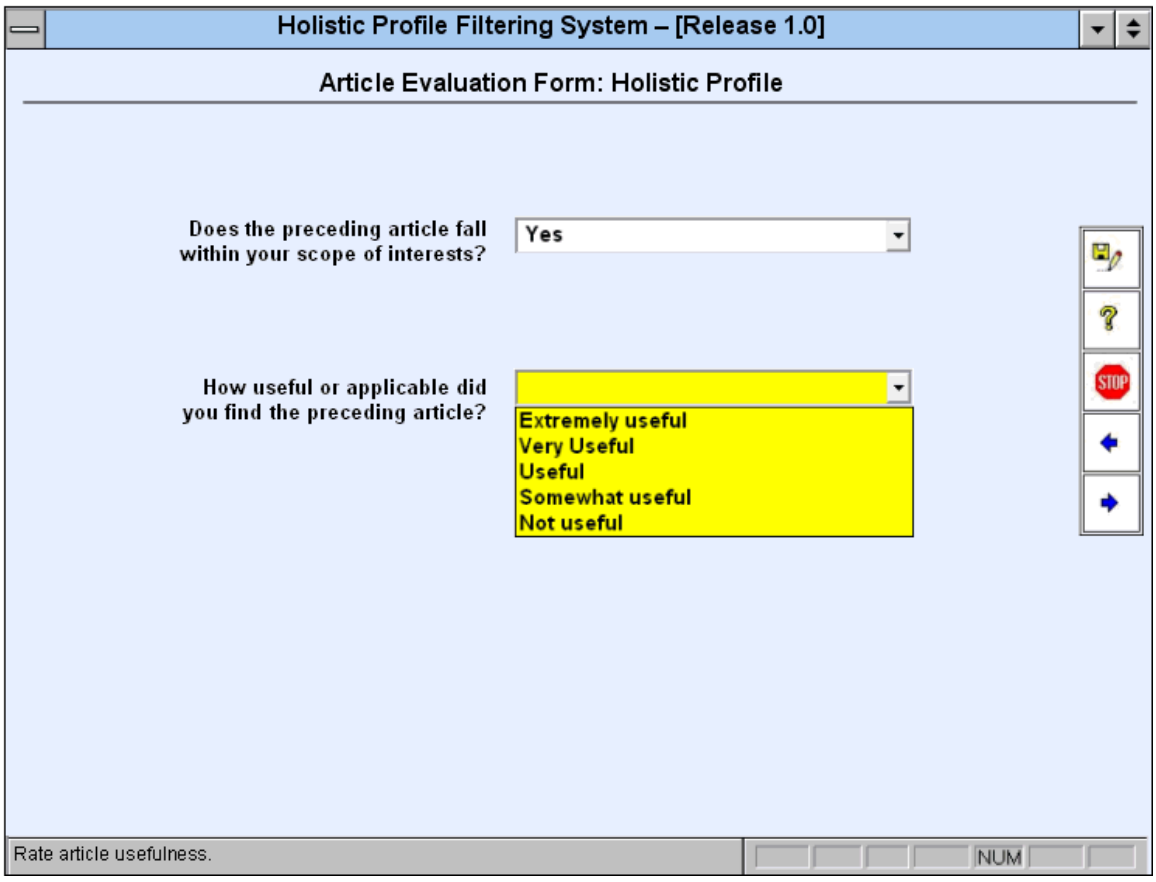

Figure A.11. Evaluation Screen.

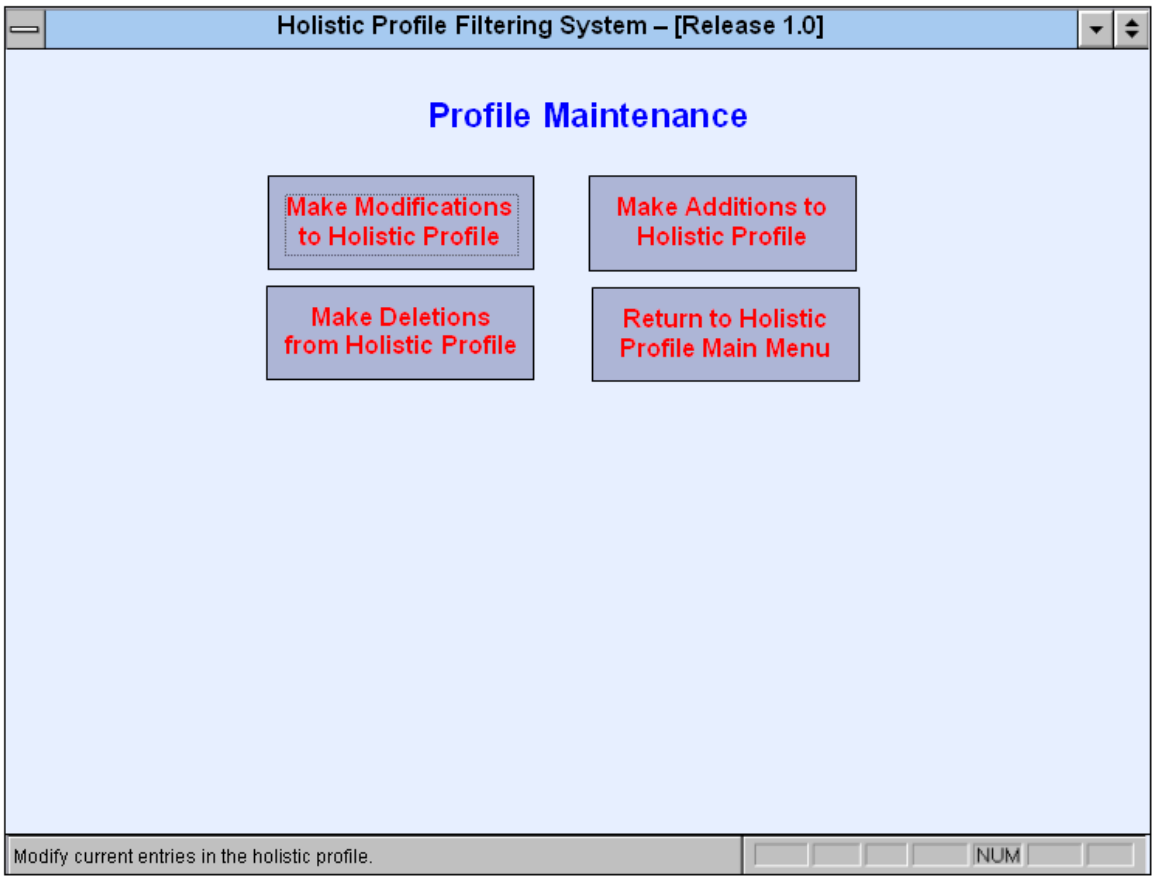

Figure A.12. Maintenance Screen.

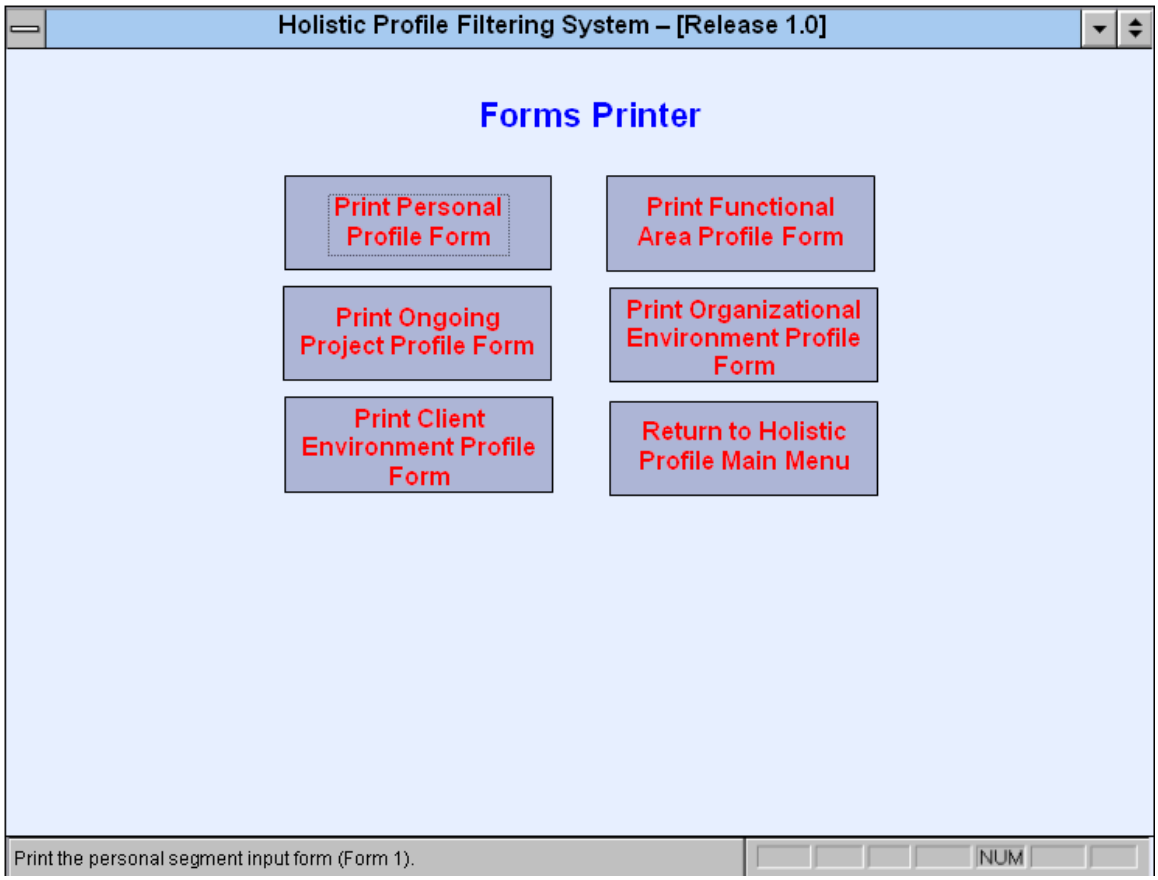

Figure A.13. Print Input Forms Screen.