Tracking Footprints through an Information Space:
Leveraging the Document Selections of Expert Problem Solvers

A. Project Summary

The goal of this project is to help an expert problem solver find needed information in a large, complex information space, locating sufficient information to solve the problem at hand without the distraction of redundant or irrelevant information. We focus on one example of expert problem solving, the health care field, where the medical record of a patient can be a large, complex, and geographically distributed collection of documents, created by a diverse array of health professionals, for divergent purposes, over an extended period of time. Sorting through such a collection, even an electronic one, to find needed information can be a formidable and time consuming task.

Our approach is to capture the trace of information used by experts as they solve problems and then to exploit this information to assist others. Imagine a heart specialist who is called in to see a patient to solve a specific problem. Reviewing the medical records of the patient, the specialist must somehow locate sufficient relevant information to understand and solve the problem, while ignoring the much larger quantity of information that belongs in the record but is irrelevant or redundant with respect to the problem at hand. As she traverses this large, diverse, often disorganized collection of documents, she makes explicit choices about which documents to ignore and which to examine more carefully. Taken together, her choices create a discrete subset of documents relevant to a given problem and likely to be of interest to subsequent users of the collection who are concerned with the same problem..

We believe that by capturing the information that is inherent in the document selection choices of expert problem solvers, we can take advantage of their knowledge (without trying to replicate their expertise) and use it to assist subsequent expert problem solvers using the collection. We propose to develop: (1) for each expert solving a problem, a trace that describes the path taken by the expert problem solver through the collection; (2) for each document, a precis of information about the document, its content, and the history of its use by experts; and (3) navigation tools that assist subsequent problem solvers using the collection by exploiting the knowledge inherent in existing traces.

The work is conducted by a cross-disciplinary team comprised of an MD whose research is focused on the information seeking behavior of physicians, and a group of computer scientists who focus on extracting and using regularly structured information. We have convened a Technical Advisory Board comprised of experts in collaborative work, users and developers of electronic medical record systems, and a specialist from Boeing Corporation. The Board evaluates and reviews the project research and serves as a mechanism for dissemination and for considering the broader impact of the work.

The usefulness of this approach is not limited to the medical domain. We expect this technology to be scalable, because the value of a set of precis and bundles within a collection should increase as the document collection grows and as more experts use the collection. We also expect the technology to be generalizable to other domains, such as aircraft development and maintenance, in which experts must navigate large, complex document collections to solve specific problems.

C. Project Description — *2*

C.1 Research Vision and Rationale — *2*

C.2 Research Objectives — *3*

C.2.1 Research Objectives: Observing Expert Information Seeking — **4**

C.2.2 Research Objectives: Developing Trace-Based Technology — **5**

C.3 Research Plan — *6*

C.3.1 Research Plan: Observing Expert Information Seeking — **6**

C.3.2 Research Plan: Developing Trace-Based Technology — **7**

C.3.3 Logistics — **10**

C.4 Dissemination and Evaluation — *11*

C.5 Relation to Present State of Knowledge in the Field — *11*

C.5.1 Related work — **11**

C.5.1.1 Present State of Knowledge: Physician Information Seeking — 11

C.5.1.2 Present State of Knowledge: Developing Trace-Based Technology — 12

C.5.2 Results of Prior Work — **13**

C.5.2.1 Prior Collaborative Research — 13

C.5.2.2 Results of Prior NSF Support — 13

C.5.3 Expected Significance — **15**

C.5.4 Relation to Longer Term Goals of PIs — **15**

C.5.5 Significance of Proposed Work and Broader Impact — **16**

C. Project Description

C.1 Research Vision and Rationale

Imagine that, while visiting a series of national parks, you could visit the places where a professional photographer stopped to take photos. At a more detailed level, perhaps you'd like to know where photographers specializing in wildlife or sunrise photography stopped. You might find this selection more interesting than trying the view from every spot in the park, on your own. In essence, we are offering the opportunity for you to track the footprints of experts in an information space and take advantage of their experience and their expertise, in the form of a trace showing their selection of information elements.[1] You can see the world from their vantage point.

We focus, in particular, on the use of traces in a problem-solving context. When detectives are trying to solve a crime, when aircraft mechanics are trying to identify and fix a problem, when physicians are trying to diagnose and treat patients, their goal is, first and foremost, to solve the problem at hand. Further, these problem solvers use a discrete, ever-increasing, historical record of events such as the body of evidence, including crime lab reports; the record of all maintenance procedures conducted and parts replaced on a given aircraft; or a patient's medical record, respectively, as one important source of information.

Our premise is that such problem solvers rarely need a comprehensive view of the historical record; only a tiny subset of information is relevant to the current problem. Further, the number of problems and their associated contexts is nearly unlimited; thus it is very difficult to anticipate ahead of time what sort of information is needed. It follows then that standard indexing techniques, e.g., used in information retrieval and databases, is not necessarily the most promising approach for identifying relevant information for problem solvers. If we consider the national park analogy, there are many guidebooks published for Yellowstone National Park, but there may not be any guidebooks for the particular location we're working with (perhaps it's a wilderness area or private property).

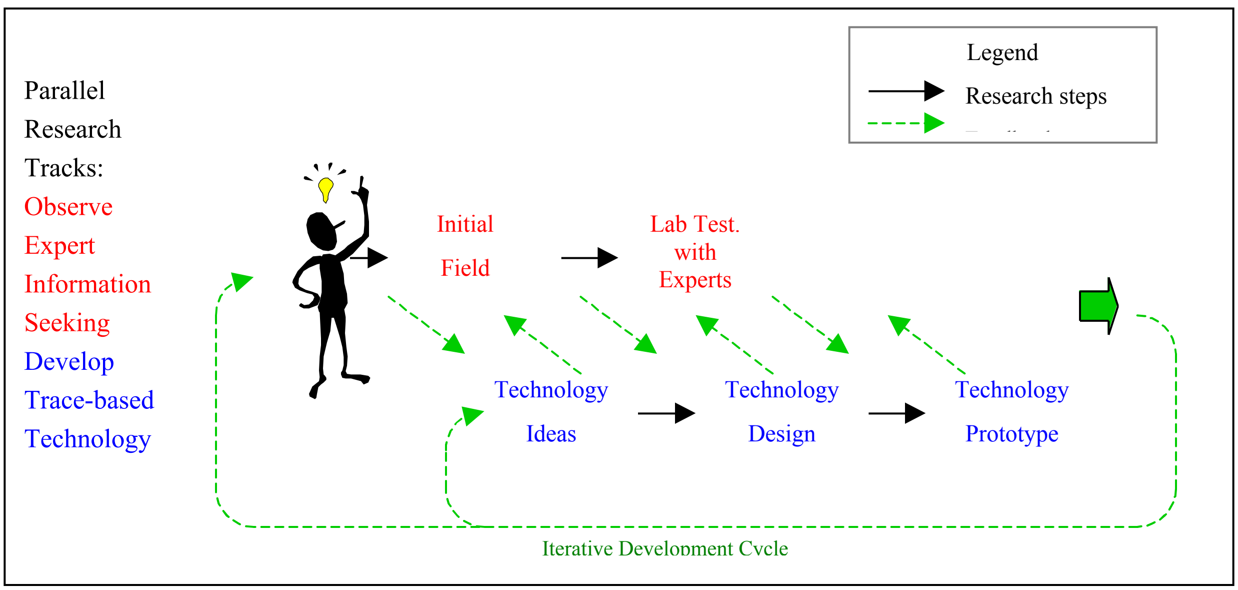

Figure 1: Work plan sketch showing two parallel research tracks, Observe Expert Information Seeking and Develop Trace-Based Technology, with feedback and iterative development

Our motivating application is a physician's use of a medical record while treating a patient. This type of information seeking is highly directed, based on the patient's current condition or complaint, the patient's medical history as reflected in the medical record, the physician's role (e.g., as an on-call physician as

[1] We use the term "information element" to indicate a useful unit of information. It could be a document, a record, etc. In practice, we require that the underlying information technology allow the user to delimit and reference information elements.

compared to a primary care physician), and the physician's expertise (e.g., cardiologist vs. a family practice physician).

We propose to track the footprints (capture the *trace* of information use/selection) of expert problem solvers. We will then exploit the expert knowledge contained implicitly in such selections by making them available to other users. As an example, an on-call physician might like to see the documents that he or she selected during the most recent encounter with this patient. A primary care physician might benefit from seeing the documents deemed relevant by a neurologist, on a recent referral. We believe that there is valuable knowledge implicit in this selection.

Our work proceeds in two parallel research tracks: Observing Expert Information Seeking and Developing Trace-Based Technology, as shown in Figure 1. The first research track, Observing Expert Information Seeking, will observe expert problem solving, develop protocols to describe the capture and use of traces, and conduct human subject experiments to determine the utility of the protocols. The second track, Developing Trace-Based Technology will develop a superimposed information structure[2] to capture and use traces.

We envision a bi-directional influence between the two tracks. On the one hand, the technology will be developed based on the results of the human subject experiments conducted in the first track to ensure that we build technology that actually helps to solve the problem. On the other hand, the objectives for the technology may "push back" on the requirements for information traces. The ultimate goal for this project is a technology that solves a recognized problem as demonstrated by human subject experiments, but that also exhibits desirable properties of being easily automated, semantically integrated, and based on structure that enables navigation and access through the superimposed information.

This work is complementary to work on search engines (where a complete information space might be indexed based on selected keywords). In a similar manner, this work is complementary to database-style query processing; problem solvers often do *not* want a comprehensive query answer. This work is also complementary to work on expert systems. We do not attempt to reproduce medical knowledge or diagnostic expertise. Rather we track the footprints of experts as they solve problems.

We have convened a Technical Advisory Board consisting of experts on collaborative work, a representative from Boeing Corporation, and both users and developers of electronic medical records systems to provide ongoing evaluation and direction for the work. The Board will actively consider the possibility of technology transfer and contribute to the potential for broad impact.

We present our research objectives and our research plan in the next two sections, Sections C.2 and C.3, respectively, with our plans for dissemination and evaluation discussed briefly in Section C.4. In Section C.5 we present background information including related work, prior work of this collaborative team, as well as results of prior NSF grants. We conclude the proposal with a brief discussion of the expected significance and broader impact of this work in Section C.6.

C.2 Research Objectives

The goal of this project is to enable an expert problem solver to efficiently navigate a large, complex information space, locate information that is necessary to solve the problem at hand while ignoring information that is not. As an example of how the technology might work, consider the mock screen layout shown in Figure 2.

In the left pane, available documents in the underlying collection can be organized by conventional attributes, such as date or indexing category, or by novel attributes, such as use history gleaned from traces. The middle pane allows detailed viewing of individual documents. Using the right pane, experts can create structured and annotated subsets of the document collection. Using a drag-and-drop method, entire documents (top arrow), selected bits of information from a single document (bottom arrows) or annotations

---

[2] We refer to "superimposed" information and a "superimposed" layer of technology. The adjective "superimposed" means that the information and technology are separate from the underlying information and technology and are introduced to supplement the underlying information and technology. The underlying information and technology continue to support their original purpose(s); the superimposed information and technology exist separately from the underlying information and technology.

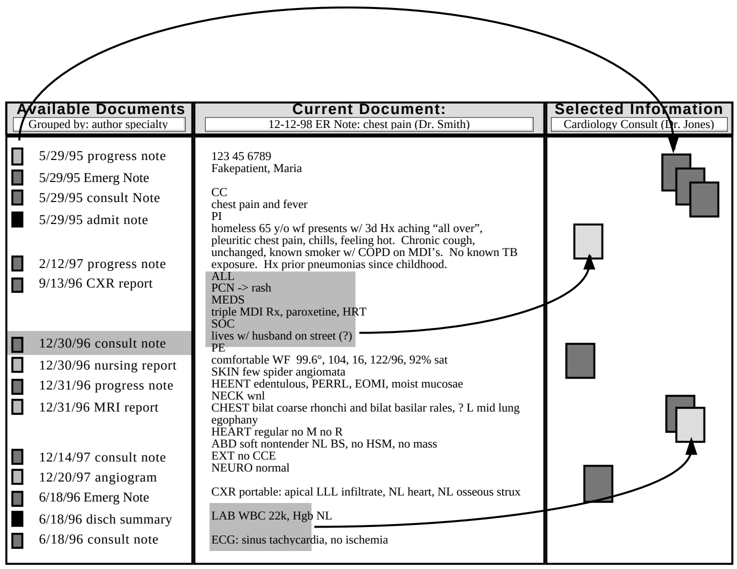

Figure 2: Three Pane Browser for Viewing, Selection, and Annotation

may be selected. Expert relevance judgments are inherent in the selection, attention (viewing time), and organization (grouping and placing within the right pane) of the information elements being examined. User actions collectively form a trace, from which these relevance judgments can be deduced.

C.2.1 Research Objectives: Observing Expert Information Seeking

The main objectives for this research track are: (1) to describe the information seeking behavior of expert physicians using a medical record to solve a clinical problem and (2) to evaluate, early and often, trace-based technology as it evolves from conceptual model through early prototype to working technology. We elaborate each of these goals with a series of objectives, posed as research questions.

Observation Research Goal 1: Describe the information seeking of expert physicians using a medical record to solve a clinical problem.

Objective 1.1. What cues or document attributes do experts use when selecting documents in a conventional, printed medical record? Experts cannot examine the content of every document in a collection. Some content- and context-independent attributes must be employed when scanning the collection to choose which documents to examine in detail. Candidate attributes might include general appearance, location in the collection, size, format, source, author, purpose, age, legibility, familiarity, expected signal-to-noise ratio, etc. What other attributes, including content, context, and history of use by previous experts, can contribute to improved document selection? An efficient trace would contain all of the attributes that are important to selection and none of the information that is not relevant to selection.

Objective 1.2. What cues or document attributes do experts use when selecting documents in an electronic medical record? Many of the attributes available with conventional records are physical attributes of the print medium, which are lost in a digital collection. How do experts navigate and choose documents from a digital collection? Are they more or less effective in this process? Does the form (print or digital) of the collection affect the subset of data to which the information seeker will be exposed and therefore potentially affect her understanding of the problem, for good or ill?

Objective 1.3. How do expert problem solvers keep track of information selected from a collection? What physical processes (bookmarks, note-taking, dog-earing corners, etc.) are employed to keep track of information chosen from a print collection? We have observed this kind of behavior when experts solve problems; experts tend to group, structure, and annotate little bits of information, e.g., on the side of a progress note, on the back of an envelope, etc. Can we provide useful alternative processes for use in searching a digital collection?

Objective 1.4. How does the problem solver know she has enough information? Is this related to the apparent size or complexity of the collection? Is this affected by the medium (digital or print)? What strategies are employed to achieve the necessary compromise between high recall retrieval and high precision retrieval?[3]

Objective 1.5. How do expert problem-solvers combine and organize information selected from a collection into new combinations for their own purposes? Does the form of the collection affect their ability to create these combinations? Does it affect the quality, completeness, or usefulness of the result (usually a novel summary document with annotation and/or interpretation)?

Observation Research Goal 2. Evaluate trace-based technology early and often, as it evolves from conceptual model through early prototype to working technology:

Objective 2.1. Is the initial conceptual framework for trace-based technology valid? Does the information-seeking behavior of expert problem solvers support this model?

Objective 2.2. Is the initial conceptual model of trace information, based on the results of 1.1 and 1.2 above, valid? Is the attribute set complete? Is it minimal? Does it translate from print to electronic document collections?

Objective 2.3. Is the initial conceptual model of trace building correct? Using a simple (even paper and pencil) prototypes, are expert problem solvers able to exploit the information in a trace to navigate the collection more effectively or efficiently? Are they more or less satisfied with its form and its use?

Objective 2.4. Once created, is the working prototype of this technology usable? Is it useful? How should it be changed to improve its usability? How should it be changed to improve its usefulness?

C.2.2 Research Objectives: Developing Trace-Based Technology

Both the nature of the information space and the nature of the information-seeking behavior exhibited by the expert problem solvers will shape the proposed technology development research.

The information space is an unpublished collection without a standard library index or catalog. The number of users of any particular piece of information is tiny, by any reasonable information retrieval or database query standards. How many doctors has any one individual consulted in his or her lifetime? The information is heterogeneous. Although paper-based systems use individual documents, an electronic medical record (EMR) system could include images (e.g., from an x-ray), videos (e.g., from an ultrasound), audio (e.g., recording of heart murmurs), etc.

The information-seeking behavior is decidedly different from standard browsing, querying, or searching. It is almost never the case that a medical record is browsed during an idle moment. Rather, a medical record is consulted when a patient has a problem and the expert is involved in a problem solving activity. The relevant access paths, e.g., based on typical queries, cannot be anticipated ahead of time, in part, because there are too many possible (medical) problems, each with its own profile of relevant information.

We have two main research goals, with associated research objectives, for the development of technology to capture traces.

Technology Research Goal 1: To limit the information in traces to information that is useful to problem solvers *and* that can be easily exploited, that is, easily extracted and manipulated.

Objective 1.1: To capture and use information that is useful to problem solvers, as determined by the Observe Expert Information Seeking research track.

Objective 1.2: To limit traces to information that can be easily extracted

Objective 1.3: To limit traces to information that can be semantically integrated in a straightforward manner. This is an essential objective because information, particularly from different sources, cannot be used together unless it has the same associated meaning.

---

[3] With high recall, every relevant document is retrieved, including redundant or irrelevant information. With high precision, every retrieved document is relevant, usually at the expense of missing some relevant and potentially decision-changing information.

Objective 1.4: To introduce only regularly structured information, to support DB style query/access.

Objective 1.5: To extract regularly structured descriptions of content from unstructured text using machine learning techniques.

Technology Research Goal 2: To support traces in a superimposed layer of information and technology, e.g., over the current medical record systems.

Objective 2.1: To minimize the interference of the superimposed layer with the underlying technology that manages the information in its original form.

Objective 2.2: To maintain the connections from an entry in the trace with the underlying, selected information element.

Objective 2.3: To include heterogeneous information types, managed by underlying, heterogeneous technologies.

Objective 2.4: To focus on access to the underlying information through the superimposed layer.

Physicians do not normally alter the existing documents in a medical record. This means that we can focus on navigation and query capability, enabled by the superimposed layer, and defer consideration of update to the underlying information elements to other research projects.

C.3 Research Plan

The research plan intertwines the activities of the Observing Expert Information Seeking research track and the Develop Trace-Based Technology track, as suggested in Figure 1 above and as described in Table 1. Activities 1 and 5 each initiate a cycle for our iterative development.

Table 1: Combined Activities for Each Development Cycle

Task	Responsibility
1. Obtain empirical data in field or lab setting	Observe Expert Info. Seek. Team
2. Report findings to combined research group	Combined Research Team
3. Revise or refine models of trace-based technology	Develop Trace-Based Tech. Team
4. Review progress with external advisory group	Combined Research Team
5. Return to field or lab setting to test revised model	Observe Expert Info. Seek Team

The detailed research plan for the two research teams are presented in Sections C.3.1 and C.3.2, respectively, followed by a brief description of our project logistics in Section C.3.3.

C.3.1 Research Plan: Observing Expert Information Seeking

The majority of the Observing Expert Information Seeking work occurs in year one and two, to obtain a sound empirical basis for constructing trace-based technology, and to allow early (and repeated) evaluation, revision, and refinement of the resulting ideas and technology prototypes.

The research focuses on the information-seeking behavior of expert medical problem solvers. To obtain valid information, eight to ten medical experts from a representative spectrum of practice domains will be recruited to participate intermittently over the 3-year project. We will also recruit several non-physicians, such as case managers, administrative personnel, or quality assurance auditors, because they must also navigate the medical record to solve a specific problem. Subjects will be recruited directly by the PI (Gorman) based on his familiarity with members of the medical staffs of each of the Portland area health care institutions. The budget includes incentives for participants, given that they are traditionally difficult to recruit. We have successfully recruited participants and conducted similar field research with over 150 physicians and other clinicians in the state [Gorman95a, Gorman97]

The main activities of this research track are listed in Table 2, as activities used to satisfy the Observation Research Goal 1. We will support the Observation Research Goal 2 (to evaluate trace-based technology early and often, as it evolves from conceptual model through early prototype to working technology) with the identical sequence and laboratory approach as shown in Table 2, with the exception that data collection will be confined to the laboratory, where the technology prototypes are most appropriately tested. [Stead94] This slightly modified evaluation cycle will be used for each iteration of the technology.

Table 2: Activities for Observation Research Goal 1: Describe information seeking of medical experts using the medical record in problem-solving.	
1. Conduct Focus Groups [Morgan97]	One or two focus groups will be held to develop an initial validation of: the conceptual model for the use of traces, an initial list of candidate attributes for inclusion in traces, and discussion of information seeking in print and electronic medical records. Focus groups are ideal for this type of initial hypothesis-generating exploration, because they allow interaction among subjects that permits cross-fertilization and generation of ideas. These initial ideas will then be tested using subsequent field and laboratory observation.
2. Conduct Participant Observation Studies	Working with each medical expert in her or his usual surroundings as they perform their usual tasks, using whatever conventional and electronic medical records are currently installed, we will observe experts in problem solving tasks, with minimal disruption or interruption. We will record behavior using conventional field notes, selected audio recording where needed and acceptable to subjects, and novel techniques such as digital writing pads to record the note-taking behavior of experts as they create traces.
3. Use Protocol Analysis [Ericsson93]	Initially, we will hold "Think Aloud" sessions using conventional print medical records. Subjects will be given scenario-based problem solving tasks, based on our early piloting of this method. Research staff will conduct ongoing observation and interviewing to identify cues and attributes, as well as record the information selections.
4. Use Critical Incident Technique [Flanagan54]	We will conduct a small number of critical incident interviews, focusing on one successful episode and one unsuccessful episode of information seeking in a conventional or electronic medical record. Critical incident technique is a well-established method for identifying factors associated with successful or unsuccessful task performance, dating back to the study of Navy pilots successful and unsuccessful missions in WWII.
5. Review with entire research team	Findings of field and laboratory studies of experts will be communicated to the Trace-Based Technology Development team in a group discussion intended to revise and refine the conceptual model, to provide the basis for the development of trace-based technology. Following these meetings, trace-based technology will proceed as described below.
6. Report to Technical Advisory Board	Findings of observation/evaluation team and refinements of technology development team will be reported to the TAB for comment, criticism, and suggestion.
7. Publish Results	We will publish results including: (1) validation of trace concepts; (2) initial specification of attributes used to describe information elements; (3) initial specification of trace from digital pad logs; (4) verification of methodology for later field studies.

C.3.2 Research Plan: Developing Trace-Based Technology

Our approach to supporting traces is based on the use of *precis*. Each precis represents one information element; it serves as a simple proxy for the underlying information element. Figure 3 presents a sketch of how we will structure the technology, with precis shown on the upper right side. The underlying information space, e.g., the medical record for a patient, is shown on the lower right side of the figure. The precis contains a few key attributes of the underlying information element, such as date that the information element was created, author of the information element, the information element type (such as lab report, physician's encounter notes). The precis is always connected to the underlying information element with something like a bookmark, plus annotation with a few descriptive attributes.

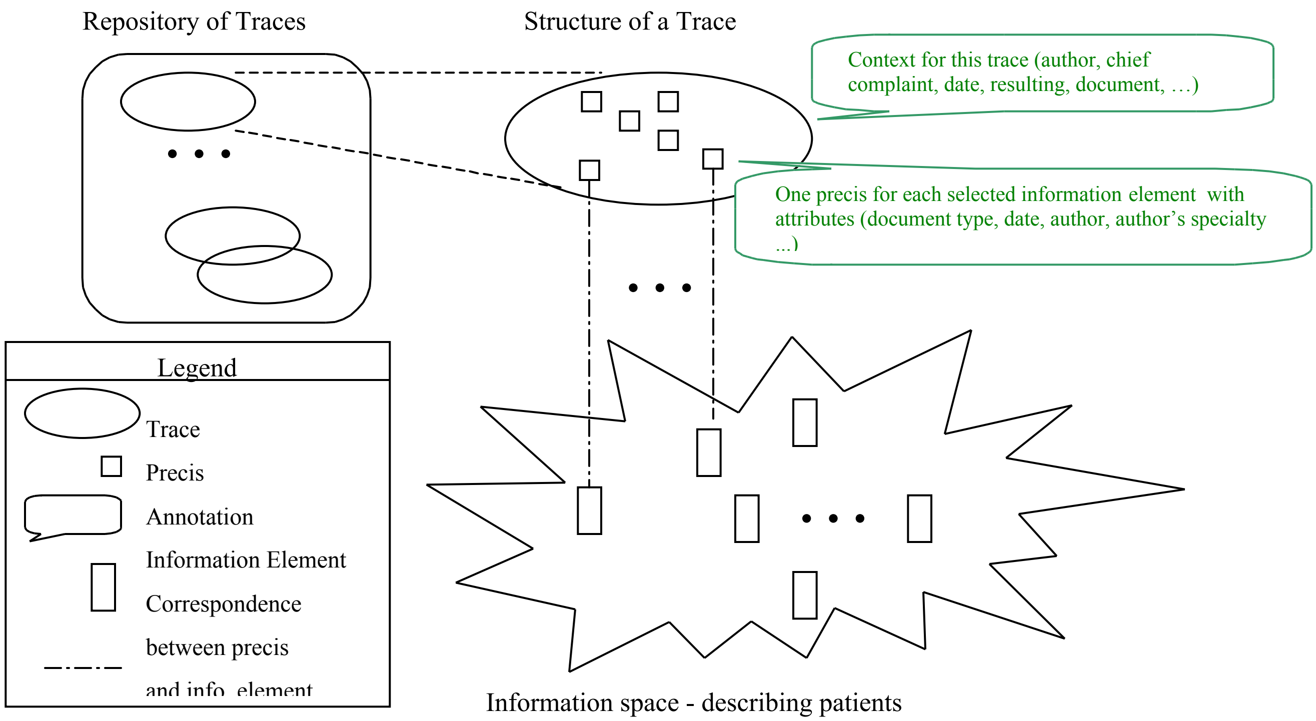

Figure 3: Representation for traces, based on precis

The use of precis distinguishes our approach to information integration from classical database and schema integration. Precis do *not* provide a complete representation for the underlying information; precis are simple, perhaps even minimalist, by design. Only attributes (of the underlying information element) that are easy to extract, straightforward to represent, and that have a well-understood meaning (e.g., MeSH codes) will be included in precis. Our use of precis also distinguishes us from various approaches to heterogeneous information integration. We provide uniform access to precis (as organized into traces) *and* we maintain the connection to the complete underlying information element, in its original context. Information seekers can easily "drill down" to the source information as needed. Precis can be created as needed; there is no need to be complete. The specific content for both precis and traces is to be determined as part of the research. The work consists of seven major activities, as listed in Table 3. These activities will be used during each iteration of the technology development.

Activity 1: The content of precis and traces, is influenced by three main forces: (1) the needs of the problem solvers, (2) the specific information and how it appears in the specific underlying information sources used for the prototype, and (3) the research goals for the technology development. This kind of tension provides a creative environment for reaching our overall goals for this project. The combined research team performs this activity, with input from the Technical Advisory Board.

Activity 2: In general, the record of any doctor-patient encounter contains a description of the patient's complaint, frequently in the patient's own words. For the reader of the record, the chief complaint provides the context in which further actions and information are understood. For a trace, chief complaint provides a description of the problem that was being addressed when the trace was created. We will attempt to develop a set of descriptors for "chief complaint" that is both informative to users and likely to be predictable from the text. Dr. Rehfuss will develop text categorization mechanisms for predicting these descriptors from the text. We expect to use (a subset of) a standard medical code sets, such as SNOMED, as these are already familiar to physicians. When choosing the codes, we must balance informativeness to users with predictability. We believe it is a novel approach to do text categorization where the set of categories is selected based (at least partly) on the predictability of its elements.

Activity 3: Once the information needs are understood, the next step is to design a suitable model and structure for the superimposed layer. Drs. Delcambre and Maier will draw on their experience in other research projects involving superimposed information [Washburn96a, Washburn96b, Delcambre97] to develop an appropriate model. This activity will be particularly influenced by Technology Research Goal 1 to have the superimposed information easy to exploit.

Activity 4: Given the model, the next step is to choose a representation for the superimposed information. This activity is influenced by both technology research goals because the choice of a representation influences the information model as well as the tools and techniques used in the implementation. We plan to use XML because it provides a standard representation for a broad range of information, it includes the ability to specify type information (analogous to a schema) that describes the data instances, it includes the ability to link to other information sources (which could enable the precis to link to the underlying source documents), and there are a growing number of tools available to process XML.

Table 3: Research Activities for Developing Trace-Based Technology	
1.	Consider information observable from an expert, during problem solving vs. the goals for the superimposed information and technology.
2.	Extract information about current problem.
3.	Develop a suitable model and structure for the superimposed information
4.	Specify a suitable representation for the information. We plan to use XML for this purpose.
5.	Develop the mechanisms for access and use of traces.
6.	Design the architecture for the trace and precis implementation.
7.	Mine the accumulated traces for “trails” and “landmarks”. (Possible activity)

Activity 5: In addition to representing individual precis and traces, we must also consider how they will be accessed. The requirements for access will come from the Observe Expert Information Seeking team and will be influenced by the details of the prototype implementation, but this task will also be strongly influenced by the opportunity to apply database-style processing.

Activity 6: The architecture and the implementation for this work will be guided by the second technology research goal, where we articulate what we mean by a superimposed technology layer.

Activity 7: Once a set of traces exists, it makes sense to look for patterns in this set of "footprints" -- "trails" and "landmarks" -- as these have a high likelihood of being relevant to multiple situations. Data mining the set of traces indirectly mines the underlying textual database. If a sufficient number of traces is gathered, Dr. Rehfuss will address these issues. In any case, implementation decisions will include this as a possibility.

The Trace-Based technology can be implemented at various levels of sophistication, in all aspects. Consider some of the options shown in Figure 4. A simple way to implement the capture of traces would be for the user to explicitly create them, e.g., by highlighting an information element and dragging it to the trace workspace. A more sophisticated implementation might capture a trace implicitly, by noticing how long an expert lingered on a certain web page or window. We also have increasingly sophisticated options for selecting a trace from existing traces. A very simple approach would support only a single trace, the one most recently created by this user. A slightly more sophisticated approach would list available traces, perhaps sorted or grouped according to useful attributes of the traces or the precis. A much more sophisticated option would be to suggest traces that might be useful for the current context. Much work can be done on tracking the relationships among traces. Can a trace be changed? Should we maintain versions of traces, much like version control systems? Are there useful relationships among traces that should be captured? What is the structure of a trace? A set? A list? A structure that reflects the way information was organized by the expert? Does a trace lead to the introduction of new information element(s) into the underlying information space, e.g., as in the physician notes for the current visit? The entire team, through the course of this project, will consider these questions.

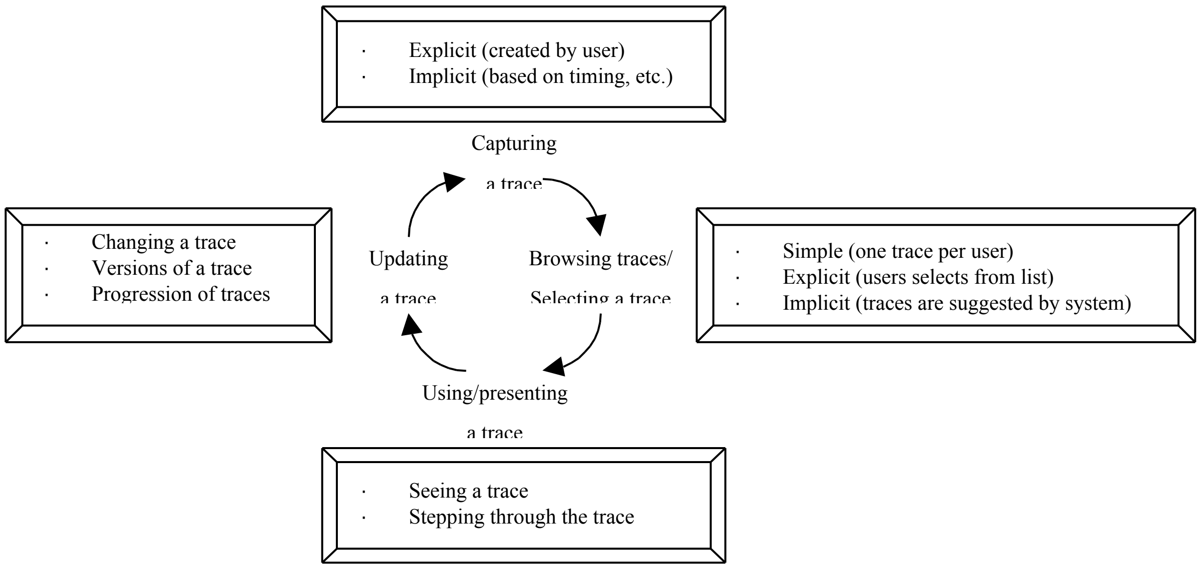

Figure 4: Increasingly sophisticated technology options to support the capture and use of traces

C.3.3 Logistics

The research team has been actively collaborating for the past eighteen months, based on funding from DOE. Our proposed project organization follows the organizational style that has been successful for us in our research to date. Drs. Gorman (OHSU) and Delcambre (OGI) will direct the parallel tracks of Observing Expert Information Seeking and Developing Trace-Based Technology, respectively. They will handle project administration and take the lead on project planning.

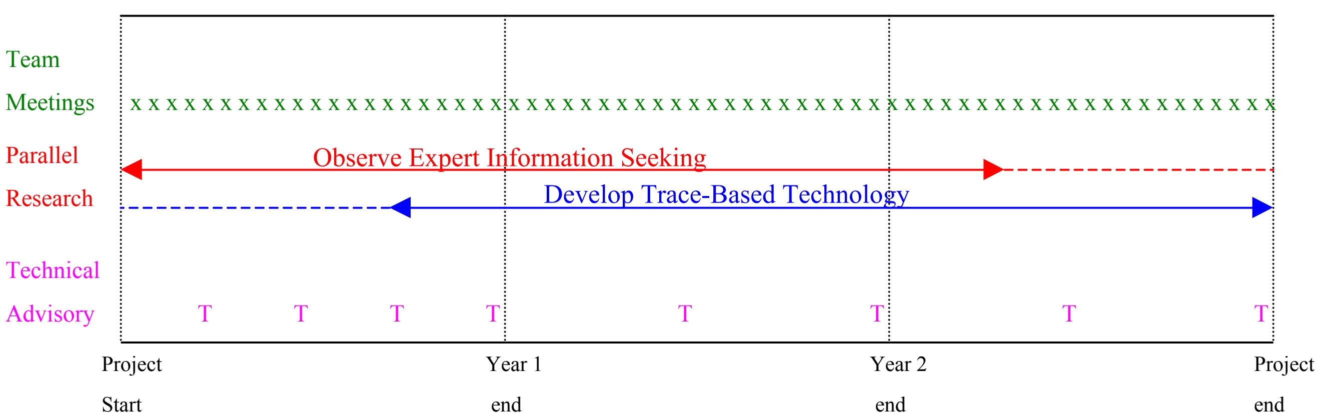

Figure 5: Overview of activities

The main progress of the work will be documented in twice-monthly team meetings, as indicated on the top line of Figure 5 with collaborative brainstorming and research also taking place at these meetings.

The second line of Figure 5 shows the two parallel research tracks, both operational throughout the three-year project. The first track, to Observe Expert Information Seeking, provides the driving force for this project. Thus this track must begin immediately upon project start and will have the highest level of activity during the first 27 months. The second track, to Develop Trace-Based Technology, follows immediately behind the first. This track will have its highest level of activity during the final 27 months of the project. This is indicated in Figure 5 by the solid and dotted lines to indicate high and medium activity levels, respectively.

A Technical Advisory Board will meet quarterly during year 1 and twice yearly during years 2 and 3 of the work, as shown in Figure 5 to maximize the opportunity for the board to influence the research direction of the project. Board meetings will be held in Portland, OR and will involve project review, brainstorming, and evaluation by the Board. Members of the Board will also guide us in applying our research within the medical domain, extending our research beyond the medical domain (to aerospace), and considering usability and collaboration issues. Members of the board are shown in Table 4.

Table 4: Technical Advisory Board Members (see letters of support in Appendix J)		
George Englebeck, PhD	Advanced Computer Technologist III	Boeing Corporation
Sara Bly, PhD	Consultant	Computer-supported collaborative work
Blackford Middleton, PhD, MD	Vice President for Clinical Informatics	MedicaLogic, Inc
Dick Gibson, PhD, MD	Medical Director, Information Systems	Providence Health System of Oregon

C.4 Dissemination and Evaluation

We have two major paths for dissemination: publishing papers and disseminating ideas, results, and possibly even technology through the Technical Advisory Board. Our plan to disseminate through publication is quite broad because of the cross-disciplinary nature of our team. We have an established record of working and publishing together and we expect to publish in multiple forums corresponding to the various disciplines that we represent. Our plan to disseminate our research results through the Technical Advisory Board reflects one of the major reasons for convening the board. In particular, the composition of the board was driven by our interest in both disseminating and evaluating our results.

Accordingly, review by the Technical Advisory Board also serves as one of our main mechanisms for evaluation of this work. As described in detail in the research plan above in Section C.3, evaluation is integrated into every cycle of the technology development process. As we iteratively discover and articulate our model of expert information seeking, it will be evaluated in the laboratory.

C.5 Relation to Present State of Knowledge in the Field

C.5.1 Related work

C.5.1.1 Present State of Knowledge: Physician Information Seeking

Previous studies of physician problem solving have provided a picture of a highly focused information seeker. Patel, comparing trainees to experienced clinicians [Patel89], and Kassirer, studying expert clincians in a history taking task [Kassirer78], found a highly focused, hypothesis testing information seeking style. Greater expertise was associated with a more direct information seeking path, acquiring <u>less</u> information to arrive at the correct diagnosis. When clinical questions arise in practice, Covell reported that most questions are never pursued [Covell85], Curley found that only the most accessible resources are utilized [Curley90], and Gorman found that physicians pursued only those questions that were most likely to have an answer, were most urgent, and were most specific to the patient-problem at hand [Gorman95a]. Increasing attention is being paid to the importance of employing observational approaches to provide an empirical basis for the design of clinical information systems and appropriate evaluation of their impact. [Forsythe91, Anderson94, Rosenal95].

C.5.1.2 Present State of Knowledge: Developing Trace-Based Technology

Related Work on Superimposed Information

Metadata is a form of superimposed information but metadata alone does not guarantee long term data viability and reuse. Metadata requires extra work that is not obviously helpful in answering an individual scientist's own research questions [e.g., Bretherton95, Bretherton97, Diederich91, Michener97].

Considerable progress has been made in the field of schema integration, even across diverse information sources [e.g., Bayardo97, Chawathe94, Genesereth97, Gravano97, Yang97]. For this work, we restrict the superimposed information to attributes with well-known meaning, we seek a structure for the superimposed information that simplifies semantic integration, and we refrain from trying to integrate complete, underlying information sources.

Much work has been done on interoperability among existing systems and databases, including work on system level interoperation [e.g., Hull97, Livny97, Wiederhold97]. We focus here on superimposing information over existing systems and we concern ourselves with retrieval-only applications, at least initially, based on the characteristics of our motivating applications of medical records and aircraft maintenance records.

Digital library efforts and novel information retrieval and distribution mechanisms for the Internet work towards finding information, particularly semi-structured [Christophides94, Buneman95, Abiteboul97] and textual information [Lynch88]. We focus on the use of traces to help a problem solver see only relevant information. We see our work as complementary to such work in the digital library field.

Related work on "Social Filtering"

Our use of traces as reusable knowledge elements is closely related to work in text filtering, known variously as *social filtering, collaborative filtering* or *recommender systems* [Avery97, Oard96]. In general, text filtering attempts to reduce information overload (as encountered in Web browsing or news reading) by selecting only a subset of the documents for the attention of the user. Bookmarks are known to be very informative ranking data for Web browsing [Rucker97].

In content-based filtering, a representation of the document content, typically based on keywords, is used for selection by matching the representation to a user profile in some way. In social filtering, what is used for selection is a rating of the document by other users, with the user being matched to the raters in some way, such as common interests. Ratings may be either *explicit*, where users explicitly enter ratings, or *implicit*, where a rating is formed from the user's behavior, for example, the amount of time spent reading a given news article. Explicit ratings have several problems: they remove all context in which the rating was generated [Procter97], they need a "critical mass" of rated documents before automatic filtering can be done [Oard96], and it is difficult to motivate users to do rating, as there is no direct benefit to them for doing so [Hill95]. Implicit ratings clearly avoid the last problem. Also, implicit ratings need to contain information about the rater [Hill95] allowing the user to decide whether to trust the rating. Information about the rater can contain contextual information, and the "critical mass" problem can be avoided by the use of a set of "trusted advisors" known to the user community to bootstrap the acquisition of ratings [Glance97]. An important point is that implicit ratings can be as good a predictor of user interest as explicit ratings. In particular this has been shown for rating based on amount of time spent reading an article [Malone87, Morita94]. A variety of data usable in forming implicit rankings is given in [Nichols97].

The proposed work is novel in several ways. The social matrix for this project involves a "tighter" community than, say a Usenet newsgroup, and the community has well-defined roles. This implies users should be better able to evaluate the rankings, and implies traces can be based on individual behavior rather than group behavior as in [Procter97]. More importantly, the proposed work involves problem-solving activity. The problem ("chief complaint") provides context and "situates" the trace [Lueg97], allowing the user to more exactly infer the relevance of the trace to the current situation. Also, we expect to be able to leverage the "note taking" behavior of physicians, generating traces that are much more information-rich, much more explicit as rankings, and much more valuable to a user, than those generated by browsing behavior. From the point of view of filtering, the precis of a document is not much different from the document itself, although whether the user "drilled down" to the underlying document may be useful in ranking.

Related work on text categorization

Much text categorization work is based on a vector-space approach to similarity [Salton97, Deerwester90] based on general characteristics of the document, such as topic or style. In contrast, our approach attempts to extract specific facts from a document [Riloff94] such as whether a given medical procedure was performed, based on the occurrence of specific (generalized) phrases, as, for example, in [Cohen96, Hersh98]. This approach is suited to terse, "note"-like documents due to their ungrammaticality, tendency toward use of stylized phrases and jargon, and high percentage of "content" words.

C.5.2 Results of Prior Work

C.5.2.1 Prior Collaborative Research

Prior work in Information Seeking Research (Gorman)

Dr. Gorman's research has focused on understanding the information needs and information seeking behavior of health professionals, including a five-year National Library of Medicine funded study "Assessment of Information Seeking in Primary Care." Results of this work have included descriptive studies of the information needs of physicians [Gorman95b], the information needs and information seeking of nurse practitioners and physician assistants, [Gorman97], and a study of the determinants of information seeking by physicians [Gorman95]. This work has been extended to include study of information management by case managers in out-of-hospital long term care of the elderly [Patterson98]. A second focus of his research has been on user-centered design of the electronic medical record [Kreis97] and evaluation of the success factors for use of physician order entry in the electronic medical record. [Ash98].

Prior Work by the Proposed Team (Delcambre, Gorman, Maier, Rehfuss)

The interdisciplinary research team for the proposed work has a record of effective collaboration. Drs. Maier, Delcambre, Rehfuss, and Gorman initially began exploring expert collaboration and problem solving as part of a Department of Energy funded effort to employ information technology to improve out-of-hospital long term care of the elderly. Results of this work included a model for evaluating the impact of information interventions in chronic disease [Gorman98] and an initial description of the precis concept as an approach to improving access to large heterogeneous collections of data in the health care domain. [Delcambre98]

Continuing this work, we have conducted "think aloud" sessions to study expert use of the electronic medical record in solving a problem in a defined clinical scenario. These studies have served not only to validate the feasibility and usefulness of this approach to studying the process, but have served to refine and refocus the underlying conceptual model. For example, we expected that clinicians would gather significant amounts of information to familiarize themselves with an unfamiliar case, but found instead that they were very highly focused, attending only to information with direct bearing on the stated problem. On the other hand, trace-building activities such as note-taking and annotation as a complement to the process of selecting and retrieving information was confirmed repeatedly in our pilot scenarios, suggesting that this activity may be an important part of collecting and reorganizing relevant information.

Prior Work on the Dictation Project (Rehfuss)

The proposed work builds on another, prior collaboration between OHSU and OGI on predicting procedure codes from dictation transcripts. That collaboration addressed issues of categorization of small sets of terse, "note"-like documents. Emphasis was on precise information extraction, rather than determining general characteristics of the document such as topic or style, fast learning methods, and dealing with highly ungrammatical text [Hersh98, Rehfuss98].

C.5.2.2 Results of Prior NSF Support

IRI-9502084: Content-Based Connections for Navigating on the NII (L. Delcambre and D. Maier 1995-96, $50,000).

This project explored the use of superimposed information models to provide new, semantically meaningful access to an underlying universe of information. The work resulted in the definition of

Structured Maps to provide domain-specific concepts and relationships over an underlying information set consisting of HTML and SGML documents with a series of prototype implementations. This project explored the use of HyTime, an ISO standard for introducing semantics over SGML documents. In particular, we were inspired by Topic Navigation Maps where topic types and topic relation types (roughly analogous to entity type and relationship types in a database schema diagram) are introduced in the superimposed layer, along with instances of both topics and topic relations. Each topic can be further elaborated with typed anchor roles. Each anchor role can hold zero or more references to information elements in the underlying universe. The key technical, HyTime construct used to define Topic Navigation Maps is the independent link (ilink). For the proposed work, we intend to use a similar approach with XML as a representation, exploiting both XML Link and XML Pointer.

SELECTED PAPERS CITING THE NSF AWARD

L. Delcambre, D. Maier, R. Reddy, L. Anderson, "Structured Maps: Modeling Explicit Semantics over a Universe of Information", Journal of Digital Libraries, Vol. 1, No. 1, 1997.

R. Reddy, L. Delcambre, "Structured Map Prototype for Superimposed Information", Poster Session, Automated Digital Libraries Forum, Washington, DC, May 1996.

L. Delcambre, C. Hamon, M. Biezunski, R. Reddy, S. Newcomb, "Dynamic development and refinement of hypermedia documents", Proc. of the EDBT '96 Conference, Avignon, France, March 1996.

R. Reddy, "Flexible mappings for retrieval from federated schemas with modeling variations", Technical Report, TR 95-015, Oregon Graduate Institute, October 1995.

BIR 96-30316: Database Support for Shared Ecological Research Sites (Database Activities Program) (N. Nadkarni, J. B. Cushing, D. Maier, L. Delcambre, J. Franklin, 1996-1997, $168,000)

A one-year, follow on activities grant for: BIR 93-07771: The Analysis of Three-Dimensional Spatial Information of Tree and Forest Canopy Structure: A Planning Activity (Database Activities Program and Ecosystems Program) (N. Nadkarni, G. Parker; 1993-1996, $132,096)

Although the ecological importance of the forest canopy has been recognized, no standard analytical methods yet exist to study the underlying structure and the spatial distribution of canopy-dwelling organisms. This grant was from the BIO Database Activities Program (DBA) to bring together forest canopy researchers, quantitative scientists, and computer scientists to work towards establishing methods to collect, store, display, analyze, and interpret three-dimensional (3-D) spatial data relating to tree crowns and forest canopies.

This proposal accomplished four specific objectives: (1) devised the foundation data for the Wind River Canopy Crane Research Facility (WRCCRF) consisting of common definitions and site-specific data to support research conducted at the site; (2) developed and evaluated a data management system for the WRCCRF; (3) supported in considerable detail two focus research studies supporting easy transfer of data among participating individual researchers; and (4) identified researchers whose work aligned in critical ways with the focus studies, so that we could consider data interconnections that would enable researchers to address questions critical to but beyond their own work.

Results were presented at invited seminars at 21 universities and sixteen presentations at scientific meetings. This research has been interpreted in the popular media, including magazines for the general public *(National Geographic World,* 1994; *Ranger Rick*, 1995; *Audubon*, 1995; *Discover*, 1995); public television (*Oregon Public Television; Bill Nye the Science Guy*, 1995); and in numerous public lectures.

Papers Citing These Two Awards

Nadkarni, N. M. 1994. Diversity of species and interactions in the upper tree canopy of forest ecosystems. American Zoologist 34:321-330.

Nadkarni, N. M. and G. G. Parker. 1994. A profile of forest canopy science and scientists − who we are, what we want to know, and obstacles we face: results of an international survey. Selbyana 15:38-50.

Nadkarni, N., G. Parker, E.D. Ford, J. Cushing, & C. Stallman. The canopy research network: a pathway for interdisciplinary exchange of scientific information on forest canopies. Northwest Science.

Nadkarni, N. M. 1995. Good-bye, Tarzan: the science of life in the treetops gets down to business. The Sciences 35:28-33.

Nychka, D. & N. Nadkarni.. Three-dimensional analysis of the distribution of epiphytes in tropical tree crowns. Biometrics.

Nadkarni, N. and J. Cushing. 1995. Final report: designing the forest canopy researcher's workbench: computer tools for the 21st century. International Canopy Network, Olympia, WA. 50 pp.

G. Spycher, J. B. Cushing, D. L. Henshaw, S. G. Stafford, N. Nadkarni, 1996. Solving Problems for Validation, Federation, and Migration of Ecological Databases. EcoInforma.

Cushing, J.B., N. Nadkarni, D. Maier, S. Knackstedt, E. Lyons, L. Delcambre, 1997. Database Support for Forest Canopy Researchers – Metadata as a byproduct of the research process. The Conference on Scientific and Technical Data Exchange and Integration, NRC.

Lyons, E., M. Sumera, S. Knackstedt, 1998. A series of three linked presentations at the annual Northwest Science conference.

C.5.3 Expected Significance

C.5.4 Relation to Longer Term Goals of PIs

The relationship of the proposed research to the longer-term goals of the investigators is summarized in Figure 6. The left-hand side of the figure shows a simplified description of our project steps. The way in which each of these activities contributes to the long-term research agendas is shown by the horizontal link to the right side of the figure. The long-term research goals of the investigators fall into three groups, discussed below.

Cycle of Project Activity:		Relation to Longer Term Research Goals of Pis
Strong empirical and experiential evidence that there's a problem and an opportunity		A trademark of each individual PI's as well as this team's research activities.
1. User-centered design & evaluation using bona-fide human experts in realistic settings	⇨	Working to ensure that physicians actually see the information they need when they need it without extra noise. (Paul Gorman)
2. Technology development, in concert with EMR vendors and users	⇨	Modeling and implementing superimposed information that supports new purposes and facilitating semantic integration. (Lois Delcambre and David Maier)
3. Technology evaluation for technical contributions & for meeting user need	⇨	Using machine learning for computation of content-based superimposed information and for determination of information possibly relevant to a situation. (Steve Rehfuss)

Figure 6: Relationship of specific project activities with the longer-term goals of the investigators.

The proposed work fits well into Dr. Gorman's long term research objectives of employing observational approaches to improve our understanding of information seeking by health professionals, using this understanding to inform the design of new information systems, and using these same methods to rigorously and comprehensively evaluate the impact of information systems on outcomes that are important to patients and clinicians.

From an information modeling and representation perspective, this project is a substantive step toward understanding what capability to reasonably and efficiently place in a superimposed layer. When compared to a conventional database management system, a superimposed layer aims to provide flexible data access capabilities to the superimposed information but *without* being the single, comprehensive repository for the information space. It's a bit like breaking the database out of the box, so that we provide database capability to a heterogeneous, federated, loosely structured set of systems (that continue functioning as intended) through the additional, structured meta-level information introduced in the superimposed layer. We believe that superimposed information spaces are increasingly important, as our access to ever-growing information spaces increases. And we believe that the notion of a superimposed information space, connected with the underlying information elements, has profound impact on the model(s) used for information as well as the architecture for the implemented system. The superimposed database management system will be expected to offer an array of database-like services yet be forced to rely on various underlying information spaces, managed by various technologies.

This proposal also addresses the *construction* of superimposed information over databases of text documents. Automating construction of the superimposed layer is necessary for economic reasons, but is problematic when the underlying database contains unstructured information, such as text or images. Given the present and foreseeable state of natural language processing, errors will occur when automatically mapping text to fixed descriptors of its content. Use of machine learning techniques for mapping text will generally associate a probability of correctness with the various descriptors. The proposed work sets up a test bed where we can begin to address questions about errors in superimposed information such as what degree of inaccuracy is acceptable and how best to make use of the probabilistic information.

From the point of view of text categorization, the proposed work continues that begun in the dictation project on information extraction from terse, “note”-like documents, based on relatively small document sets. It also addresses the novel situation where there is at least some freedom to choose the set of target categories on the basis of their predictability.

Finally, learning “trails” and “landmarks”, that is, discovering patterns in the set of traces, is an example of mining textual databases. We expect the proposed work will to give some insight into the problems of mining unstructured and errorful data.

C.5.5 Significance of Proposed Work and Broader Impact

Digitization of a document collection such as the electronic medical record has many potential advantages, but can create problems as well. Users may have difficulty navigating the collection as it grows intractably large and seemingly uniform. The technology we propose to develop could provide significant assistance to health professionals as they search these expanding information spaces, hoping to locate relevant information efficiently without missing crucial data or being inundated with irrelevant or redundant information. As electronic medical records are more widely employed and cues to navigation in print media are no longer available, our technology could help bridge the transition to a digital information space by providing new means of choosing, annotating, and organizing information.

From a knowledge management point of view, this work is generic. We make no attempt to reproduce human expertise; rather we simply acknowledge it and attempt to exploit it in a passive, non-intrusive way. This project may suggest a range of new questions to investigate based on the choice of information represented in the superimposed layer, the manner in which it is presented for use by problem solvers, and the sophistication of the knowledge used to manage it.

From a technology point of view, the most significant impact of this work is the exploration of a new paradigm for information access that does not require all information to be brought into a single, monolithic, managed environment (like a traditional database management system). Our technology requires only the ability to access and remain connected to underlying information elements of interest. This paradigm is complementary and orthogonal to various, underlying information management paradigms such as database management systems or text-retrieval systems. Once the basic concept of superimposed information is demonstrated, there will be much to explore based on the level of sophistication of the underlying technology and the potential for cooperation between the superimposed and the underlying technologies.

We also believe this technology will extend readily to expert problem solvers in other domains. In aerospace engineering, to name one example, this technology may be useful in exploring the maintenance records during the life of an aircraft, maintaining complex long term relationships between manufacturer and customers, and navigation through information on an intranet by aerospace engineers as they seek information to solve specific problems, as discussed by our Technical Advisory Board member from Boeing Corporation.

References

[Abiteboul97] Abiteboul, S., Quass, D., McHugh, J., Widom, J., and Wiener, J., The lorel query language for semistructured data, Journal of Digital Libraries, 1(1), 1997.

[Andersen94] Andersen JG, Aydin CE, Jay SJ. Evaluating healthcare information systems: methods and applications. Thousand Oaks, CA: Sage , 1994.

[Ash98] Ash JS, Gorman PN, Hersh WR. Physician order entry in US Hospitals [submitted, 1998 AMIA Fall Symposium].

[Avery97] Avery, C. and Zeckhauser, R., "Recommender systems for evaluating computer messages, Communications of the ACM, 40(3), 88-89, 1997.

[Bayardo97] Bayardo, R.J. Jr., et al., InfoSleuth: Agent-Based Semantic Integration of Information in Open Dynamic Environments, Proceedings ACM SIGMOD, 1997, pp. 195-206.

[Bretherton94] Bretherton, F.P. and Singley, P.T., Metadata: a user's view, Seventh International Working conference on Scientific and Statistical Database Management (SSDBM), J.C. French (ed.), 1994, IEEE Computer Science Press.

[Bretherton97] Bretherton, F.P. and Hibbard, W.L., Metadata: A case study from the environmental sciences, Proceedings of the Ninth International Conference on Scientific and Statistical Database Management (SSDBM), D. Hansen and Y. Ioannidis (eds.), 1997, IEEE Computer Science Press.

[Buneman95] Buneman, P., Davidson, S., and Suciu, D., Programming constructed for unstructured data. In Proceedings of the Fifth International Workshop on Database Programming Languages, Buggio, Italy, September 1995.

[Chawathe94] Chawathe, S.H., Garcia-Molina, H., et al., The TSIMMIS Project: Integration of Heterogeneous Information Sources, Proceedings of the 10th Anniversary Meeting of the Information Processing Society of Japan, 1994, pp. 7-18.

[Chrisophides94] Christophides, V., Abiteboul, S., Cluet, S., and Scholl, M., From structured documents to novel query facilities, in Proceedings of the ACM SIGMOD Conference, Minneapolis, MN, May 1997.

[Cohen96] Cohen, W. and Y. Singer. *Context-sensitive learning methods for text categorization*. in *Proceedings of the 19th Annual International ACM Special Interest Group in Information Retrieval*. 1996. Zurich,

[Covell85] Covell D, Uman G, Manning P. Information needs in office practice. Are they being met? Ann Intern Med 1985;103:596-599.

[Curley90] Curley S, Connelly D, Rich E. Physicians' use of medical knowledge resources: Preliminary theoretical framework and findings. Med Dec Making 1990;10:231-241.

[Deerwester90] Deerwester, S., *et al.*, *Indexing by latent semantic analysis.* Journal of the American Society for Information Science, 1990. 41:391-407.

[Delcambre97] Delcambre, L., Maier, D., Reddy, R., and Anderson, L., Structured Maps: modeling explicit semantics over a universe of information, Internation Journal of Digital Libraries, 1(1), 1997, pp. 20-35.

[Delcambre98] Delcambre L, Gorman P, Maier D, Reddy R, Rehfuss S. Précis-based navigation for familiarization. MEDINFO 98 [CDROM edition]: 9th World Congress on Medical Informatics, Seoul, Korea, August 18-22, 1998

[Diedrich91] Diedrich, J. and Milton, J., Creating domain specific metadata for scientific data and knowledge bases, IEEE Transaction on Knowledge and Data Engineering, 3:4, 1991, pp. 421-434.

[Ericsson93] Ericsson KA, and Simon HA. "Protocol Analysis." 1993. Cambridge: MIT Press.

[Flanagan54] Flanagan JC. "The Critical Incident Technique" Psychol. Bull. 1954;1:327-358.

[Forsythe91] Forsythe DE. Buchanan BG. Broadening our approach to evaluating medical information systems. Proceedings - the Annual Symposium on Computer Applications in Medical Care. :8-12, 1991.

[Genesereth97] Genesereth, M., Keller, A., and Duschka, O., Infomaster: An Information Integration System, Proceedings of the ACM SIGMOD Conference, 1997, pp. 539-542.

[Glance97] Glance, H., Arregui, D., and Dardenne, M., "Knowledge Pump: Community-centered Collaborative Filtering", Proceedings of the Delos Workshop on Collaborative Filtering, Budapest, November 1997.

[Gorman93] Gorman PN. Does the medical literature contain the evidence to answer the questions of primary care physicians? Preliminary findings of a study. Proceedings of the 17th Symposium on Computer Applications in Medical Care; Nov 1-3, 1993. Washington, DC. New York,: McGraw-Hill. pp571-575

[Gorman94] Gorman PN, Wykoff L, Ash J. Can primary care physicians questions be answered by using the medical journal literature? Bull Med Libr Assoc 1994;82:140-146.

[Gorman95a] Gorman PN, Helfand M. Information seeking in primary care: How physicians choose which clinical questions to pursue and which to leave unanswered. Med Dec Making 1995;15:113-119

[Gorman95b] Gorman PN. Information needs of physicians. J Amer Soc Inform Sci. 1995;46:729-736.

[Gorman97] Gorman PN, Ash J, Poulsen S. "Information Needs of Nurse Practitioners and Physician Assistants" Medical Library Association Annual Meeting; Seattle, WA; May 27, 1997

[Gorman98] Gorman PN. Information Interventions in Chronic Disease. MEDINFO 98 [CDROM edition]: 9th World Congress on Medical Informatics, Seoul, Korea, August 18-22, 1998.

[Gravano97] Gravano, L., Chang, C.K., Garcia-Molina, H., STARTS: Stanford Proposal for Internet Meta-Searching, Proceedings of the ACM SIGMOD Conference, 1997, pp. 207-218.

[Hersh98] Hersh, W., Leen, T., Rehfuss, S., and Malveau, S., "Automatic prediction of trauma registry procedure codes from emergency room dictations". *MEDINFO 98*. 1997. Seoul, South Korea

[Hill95] Hill, W., Stead, L., Rosenstein, M. and Furnas, G. "Recommending and Evaluating Choices in a Virtual Community of Use. Proceedings of CHI'95, ACM Press, Denver, May 1995.

[Hull97] Hull, R., Managing Semantic Heterogeneity in Databases: A Theoretical Perspective, Proceedings of the Symposium on Principles of Database Systems (PODS), 1997, pp. 51-60.

[Kassirer78] Kassirer JP, Gorry GA. Clinical problem solving: a behavioral analysis. Annals Internal Medicine 1978;89(2):245-255.

[Konstan97] Konstan, J.A.., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon L.R. and Riedl, J., "Applying collaborative filtering to Usenet news, Communications of the ACM, 40(3), 77-87, 1997.

[Kreis97] Kreis C, Gorman PN. Word Frequency Analysis of Dictated Clinical Data: A User-Centered Approach to the Design of a Structured Data Entry Interface. Proceedings of the 1997 American Medical Informatics Association Fall Symposium; October 25-29, 1997. Nashville, TN. New York,: McGraw-Hill. pp724-728.

[Livny97] Livny, M., et al., DEVise: Integrated Querying and Visual Exploration of Large Datasets, Proceedings of the ACM SIGMOD Conference, 1997, pp. 301-312.

[Lueg97] Lueg, C., "Social Filtering and Social Reality", Proceedings of the Delos Workshop on Collaborative Filtering, Budapest, November 1997.

[Lynch88] Lynch, C., and Stonebraker, M., Extending user-defined indexing with application to textual databases, in Proceedings of the 14th VLDB Conference, Los Angeles, CA, August 1988.

[Michener97] Michener, W.J., Brunt, J., Helly, T., Kirchner, and Stafford, S., Nonspatial Metadata for the Ecological Sciences, Ecological Applications, 7(1), 1997, pp. 330-342.

[Morgan97] Morgan, D. "Focus Groups as Qualitative Research"

[Morita94] Morita, M. and Shinoda, Y., "Information filtering based on user behaviour analysis and best match text retrieval, Proceedings of the 17th ACM Annual International Conference on Research and Development in Information Retrieval (SIGIR'94), Dublin, Ireland, Springer-Verlag, 272-88, 1994.

[Nichols97] Nichols, D. M., "Implicit Rating and Filtering", Proceedings of the Delos Workshop on Collaborative Filtering, Budapest, November, 1997.

[Oard96] Oard, D.W., and Marchionini, G., "A Conceptual Framework for Text Filtering", Technical Report CAR-TR-830, Juman Computer Interaction Laboratory, University of Maryland at College Park, 1996.

[Patel89] Patel V, Evans D, Groen G. Biomedical Knowledge and Clinical Reasoning. In: Evans D, Patel V, eds. Cognitive Science in Medicine. Biomedical Modeling. Boston: MIT Press, 1989:53-112.

[Patterson] Patterson PK, Gorman PN, Caton K. Information Needs of Case Managers [submitted, J Amer Med Inf Assoc]

[Procter97] R. Procter and A. McKinley, "Social Affordances and Implicit Ratings for Social Filtering on the Web", Proceedings of the Delos Workshop on Collaborative Filtering, Budapest, November 1997.

[Rosenal95] Rosenal TW. Forsythe DE. Musen MA. Seiver A Support for information management in critical care: a new approach to identify needs. Proceedings - the Annual Symposium on Computer Applications in Medical Care :2-6, 1995

[Rehfuss98] Rehfuss, S. and Leen, T., "Detecting Concepts in Medical Dictations," submitted to Neural Information Processing Systems, 1998.

[Riloff94] Riloff, E. and Lehnert, W., "Information Extraction as a Basis for High-Precision Text Classification", ACM TOIS, 12(3), 1994, 296-333.

[Rucker97] Rucker, J. and Polanco, M., "Personalized Navigation for the Web. Communications of the ACM, 73-75, March 1997.

[Salton97] Salton, G. and Buckley, C., "Term-weighting approaches in automatic text retrieval". In K. Sparck Jones and P. Willet, editors, Readings in Information Retrieval, Morgan Kaufmann, 1997.

[Stead94] Stead WW. Haynes RB. Fuller S., et al. "Designing medical informatics research and library--resource projects to increase what is learned." J. Amer Med Informatics Assoc. 1(1):28-33, 1994.

[Washburn96a] Washburn, G., Delcambre, L., and Whiting, M., Queries from outer space, in Proceedings of the 1996 IEEE Research Issues in Data Engineering (RIDE) Workshop, New Orleans, La., February 1996.

[Washburn96b] Washburn, G., Delcambre, and Whiting, M., Reuse of high-level information requests: Leveraging the investment, in Proceedings of the Eighth International Conference on Scientific and Statistical Database Management, Stockholm, Sweden, June 1996.

[Wiederhold95] Wiederhold, G., and Genesereth, M., Basis for Mediation (updated in 1997), Proceedings of the Conference on Cooperative Information Systems (COOPIS '95), Vienna, Austria, pp. 138,155.