HEWLETT PACKARD

Multiple Views on Filtering: Consumers, Producers and Filtering Technology Designers

Vicki O'Day, Lucy Berlin; Robin Jeffries
Software and Systems Laboratory
HPL-91-153
October, 1991

information filtering

We discuss the implications of information consumer and producer needs for the provider of basic filtering technology. To understand the consumer point of view, we outline seven different filtering goals that cover a wide variety of filtering questions, including obtaining an overview, identifying a trend, and finding a match between a prototype and an instantiation. We discuss three problems from the producer point of view: structuring information, disseminating information and selecting a medium in which to express information. Both the consumer and producer perspectives are illustrated by examples from the domains of clinical patient care, medical research, and industrial product development.

Internal Accession Date Only

1 Introduction

In information-rich work domains such as medical research, clinical patient care and industrial product development, large amounts of information are generated and used for a variety of complex tasks. For a particular professional activity, only some of the enormous amount of available information is useful. In fact, access to too much information can make an activity more difficult to manage by obscuring the part of the data that is directly relevant. It is crucial for information systems in settings like these to support filtering. We take a broad perspective on filtering and use the term to include any techniques that involve reducing information for practical use.

In constructing a useful view of selected information from a large information space, two points of view should be considered: the information consumer and information producer. The information consumer, usually with a question or problem in mind, explores a large information space by requesting filtered views and interpreting the results when they appear. The information producer is responsible for the structure chosen to describe the information, what medium it is expressed in and with whom it is shared. Information consumers and producers in many settings are drawn from the same group of people; the same individual might record and disseminate information at one point and try to locate and filter information at another.

The capabilities of a system's information infrastructure contribute to the effectiveness of both consumers and producers. What are the methods available to information consumers for expressing filtering requests? What display techniques are available for viewing the results? Can consumers refine filtering requests? Can the system store and display multimedia information, or must the producer translate non-textual information into text? What dissemination options are available to the producer of new information?

The design of an information infrastructure, including its support for filtering, provides the answers to these questions through its functionality and flexibility. At one extreme, the system might offer users a collection of tools with which to design and implement their own filtering techniques. At the other extreme, the system might provide a set of templates for storing information and a fixed set of filtered views for displaying it. Clearly, there are many possibilities between these two extremes. Whatever the approach, the infrastructure provider must consider the needs of people with each of the perspectives outlined above.

It is helpful to examine these perspectives in the context of specific application domains. This exposes the variety and complexity of the problems and choices faced by workers in these domains, which in turn can motivate design choices for filtering technology. We use examples from medicine and product development to explore the consumer and producer points of view and their implications for the provider of filtering infrastructure.

Note that in some domains, such as medicine, skilled professionals who are trained in specialized information storage or search techniques in a particular field act as intermediaries between consumers and producers. An intermediary, if one exists, maps requests from a consumer into an existing information space, clarifying and focusing both the request and the results. An intermediary may also contribute to the structuring of the information; in medical literature, for example, indexing is done by intermediaries rather than authors. In this paper we will not discuss the role of the intermediary, though we believe it has interesting aspects that are not simply extensions of the producer and consumer perspectives.

The rest of this paper is organized as follows. First we present some dimensions of filtering that are pertinent to the infrastructure provider and/or filter designer. Next we describe filtering from the consumer's point of view, outlining seven fundamentally different filtering goals that cover a wide variety of filtering questions; each goal suggests the application of different filtering techniques. We then outline the filtering issues raised by the information producer's need to structure, disseminate and possibly transform information.

2 Dimensions of Filtering

In this section we describe some dimensions of filtering that can help filter designers or information infrastructure providers think about their design choices. We will return to these dimensions in the detailed discussions of the consumer and producer perspectives.

Filtering is not an exact procedure. A filtering operation is one that causes some information to emerge from a larger set, to be used by the consumer for a particular question, task or interest. Using this broad definition of filtering, the distinctions among database querying, information retrieval and filtering are not always clear. Our perspective is that the activity of filtering includes the user's efforts both to specify the filter and to apply the resulting information to his or her original question. Filtering is an *interpretive* process, which makes it less direct than performing simple database queries. Querying and information retrieval techniques can be used as part of that activity, but they do not cover the whole picture.

Browsing can be seen as a variant of filtering. It is exploratory, as filtering activities often are. When users browse, they follow a path to visit or view interesting pieces of related information from among a variety of possible choices. The accumulation of visited items can be seen as a filter over the information space. The browsing path itself is a time-based partitioned view of the filtering result set. For some filtering goals, this kind of step-wise approach may be the most effective.

Filtering can be *retrieval-based*, *presentation-based*, or both. In retrieval-based filtering, the filter is used to select which information from the data source should be fetched and presented and which should be omitted. A database query can be the instrument of a retrieval-based filter. In presentation-based filtering, display techniques are used to encourage the user to focus on some information at the expense of other available information. Generalized fisheye views [Fur86] are an example of presentation-based filters, since they encourage the user to look at nearby and important objects by increasing the salience of those objects on the screen.

Filtering can be *interactive* or *non-interactive*. An interactive filter permits the user to adjust his or her point of view and see a different subset of information emerge. A non-interactive filter requires the user to reformulate the filter specification from scratch in order to see a new subset. Non-interactive filters do not imply that the interface itself is not interactive, rather that the construction of the filter is done once, not repeatedly. The more exploratory the user's task is, the more useful an interactive filter is likely to be. Traditional queries support non-interactive filtering; query-by-reformulation [Wil84, WTF+89] and display techniques such as panning and zooming support interactive filtering.

3 Information Consumer

To apply effective filtering techniques, it is necessary to understand the goals of the person who will be using the information. Different tasks or goals should lead to substantially different filtering techniques. For example, consider the variety of questions a physician might have in approaching a clinical information system:

- What is this patient's history of vision loss?
- What is the relationship between a recent set of symptoms and a recent course of drug therapy?
- How should these test results be interpreted?
- Why was this lab test ordered?

To answer each of these questions, the physician accesses the same set of data, the patient record. However, each question has very different implications for how to select and present a subset of the information that will give the physician the clearest view towards finding an answer.

Of course, the clearest and most straightforward approach to answering a question may reduce the user's chances of making useful but unexpected discoveries through browsing. This is the potential disadvantage of tailoring filters to the consumer's primary task. In evaluating the application of different filtering techniques, we should try to understand the extent to which they introduce blind spots or remove information access points that are important to consumers.

In the domains of medical research, clinical care, and industrial product development, we identified a set of typical information filtering questions asked by professionals during the ordinary course of their work. Our sources for these work practice data include an ethnographic study of information needs in clinical practice [Faf90, FYT91], a study of usage patterns of the National Library of Medicine's online medical literature service [WSSC89], and our own interviews and observations of product development groups in an industrial setting. We examined these typical questions and found that they correspond to seven fundamentally different filtering goals:

- Obtaining an overview
- Identifying a trend or connected thread
- Learning the rationale for a decision
- Linking or mapping between multiple views
- Finding a match between a prototype and an instantiation
- Making a personalized selection from broadcast data
- Focusing on one or more distinct topics in an information space, without losing touch with each topic's surrounding context

This list may not be exhaustive, but it seems to cover a wide variety of filtering questions. In this section, we outline the seven goals, give examples of each goal, and explore the implications each goal has for filtering infrastructure.

3.1 Overview

Overview is a broad term that has several distinct variants, including ***summary***, ***transformation***, ***representative selection*** and ***abstraction***. In general, obtaining an overview implies coalescing similar or related information to produce a smaller collection of unique data points.

A summary provides a characterization of the common features of a collection of data. This is one type of transformation; other types might involve mapping data from a "raw" form into a form that carries more useful semantics, possibly losing information along the way. A representative selection can be created by labeling a cluster of data points with a typical example from within the cluster. Finally, an abstraction provides a generalization that can be used in place of a cluster of data points. In a sense, the other types of overview "look down" reflecting certain attributes of data points in the collection, while an abstraction "looks up", capturing some essential behavior that can be used by higher-level operations.

Example: *What areas of medical research were being funded this year?*

Example: *Is this patient compliant?*

An overview is not a simple subset of information. A central issue for designers of overview filters is how they are created from the raw data. Clearly, an overview may have an entirely different representation from the data, and might even be expressed in a different medium. Many overviews, such as the one required to answer the question about a patient's typical compliance, require some access to domain semantics. Other overviews, such as the answer to the funding question, could be created using structural properties of the data, such as the indexing scheme for medical literature. Abstractions depend on both domain semantics and on the point of view of the recipient. The same underlying data might support more than one abstraction, depending on the context of the information's use.

In some cases, the construction of an overview occurs mostly at retrieval time, as the data is fetched and massaged. An example is when raw instrument data (such as that collected by a chemical analysis instrument) is transformed mathematically into information with higher-level semantics (for instance, to permit the discovery of patterns). However, most overviews also rely heavily on presentation techniques to enhance the user's understanding. Display choices can help representative selections stand out or provide visualizations of clusters or shapes within the overview.

Overviews seem less interactive than some other filtered views; once an overview has been created to answer a particular question, it is unlikely to be recalculated unless the user asks a different question.

3.2 Trend or connected thread

Identifying a trend or connected thread of information is a common filtering goal. It is useful when examining time-based data, such as the history of a patient's condition.

Example: *What is the history of vision loss for this person?*

Example: *What is the market trend for improvements in the color quality of printers?*

A trend may refer to historical data, such as information about past vision loss, or it may project into the future by extrapolating from current data. In either case, the filtering goal is to show the same or related data at different points in time in order to learn something about the process of change. Trends are examined for a variety of reasons, such as to understand which factors contribute most heavily to the trend, what the speed of change is, or whether the speed of change is itself changing over time.

Trend filters are both retrieval- and presentation-based. In most of our example trend questions, a clear indication of the data to be retrieved was embodied in the question, though domain semantics might be necessary to identify all the related data. The data must then be presented in such a way that the trend can be tracked by the user. Presentation choices are critical, since some methods (such as tables of numbers) can obscure trends. Graphical methods are especially effective for depicting trends since a large amount of information can be presented in one view, thus allowing the visual comparison among data points that is necessary for this task [Tuf83]. The Perspective Wall visualization [MRC91] has been created for time-based data; it might be useful for detecting trends, especially if the table-based format used within each panel of the wall could be generalized to other visual forms.

3.3 Rationale: answer to a "why" question

A rationale is another common time-based filtering question. Here, the user wants to understand the context that led to a particular decision. The task of the filter designer is to identify and present the most relevant contextual information.

Example: *Why was this lab test ordered?*

Example: *Why was this design decision made?*

The rules for determining rationale must be present in the application domain; a filter designer cannot derive them. They might be available as a static rule collection, such as a set of "design for manufacturability" heuristics which are attached to design decisions as they are made. Or the rationale might be covered by a set of links between criteria and actual choices, kept up-to-date by decision-makers.

In the examples we saw, the people likely to ask rationale questions were professionals who understood the decision-making process involved. This implies that the filter designer should focus on presenting the contextual data itself, rather than instructing the user on how the context was used. The context will necessarily be incomplete, since only a portion of the relevant contextual information will be online in most cases. However, even partial context can help users with domain expertise infer the answers to their "why" questions.

This kind of filter might be a very interactive one, depending on how deeply the user wants to understand the rationale. A step-wise approach might be used to allow the user to probe beyond the initial context.

3.4 Linking between multiple views

Sometimes a user asks filtering questions to detect relationships between different collections of data. These are detective-like questions, covering relationships not expressed in terms of rationale and not necessarily known at the time the information was captured.

Example: *What do these different lab tests taken at the same time show about the patient's condition?*

Example: *What is the relationship between these recent symptoms and the current course of drug therapy?*

Example: *How do product defect rates relate to manufacturing process changes?*

This kind of filtering is exploratory and interactive. Though some retrieval filtering is used to include only information from the target areas (e.g. symptoms and drug therapy), the exact components that will turn out to be relevant are not known at retrieval time. Much of the functionality of this filtering is in the power of the presentation. It must give the user the ability to try out hypotheses about relationships, focusing on different parts of the data during the process. Both the user and the system must have a model of the semantic overlap between the two views. For example, the recent symptom and drug therapy views are both organized along a time dimension and relate to the same underlying biological system (the patient). The work done by MacDonald and Stuetzle on understanding multi-dimensional data through "painting" between multiple views [MSB90] is tailored to this filtering problem.

3.5 Matching between prototype and instantiation

In many domains there are formalisms on which practitioners model their processes or artifacts. In medicine, for example, the symptoms known to be associated with a disease are a prototypical set; individual patients may have some or most of them, and the art of diagnosis relies on how effectively the physician can match the actual symptoms against the prototypical set. There are many other examples in both medicine and product development.

Example: *How do I interpret these test results for this patient?*

Example: *Which methodology is most appropriate for my study?*

Example: *Which hardware or software module should I choose from this library of reusable components?*

The activity of comparing an instantiation (such as a set of lab test results for a particular patient) against a prototype (such as the description of what the results are expected to look like in general) is an activity that involves testing hypotheses. The user is looking for a best match for a particular instantiation from among the formalisms that are available. The filter designer must support easy comparison between the two, perhaps by using the same representation for both the formalism and the instantiation. A transformation could be done on one or the other as part of the filtering process, or the filter might select an appropriate representation from a set of equivalent choices.

Even if the comparison between the prototype and instantiation is straightforward, this filtering activity might be well-supported by a navigation capability, since the user might want to pursue a path some distance and then put it aside to explore another path from a branch point. It is possible that the user will not be able to decide which is the best fit unless more than one path has been explored. History mechanisms used for navigation through hypertexts [UY89] could be applied here. If the filter designer has access to domain semantics that encode the notion of matches and can compare matches to one another, the system can explore multiple paths in parallel and present the resulting best match to the user.

3.6 Personalized selection from broadcast data

This filtering goal is less specifically task- or problem-oriented than the others we outline here. When there is a body of continually growing or changing information, users often need to be notified of information that is relevant because of their current interests or areas of responsibility.

Example: *What is the latest literature relevant to my specialty?*

Example: *What is a snapshot of the current state of the automated assembly line?*

Example: *Which current design changes affect my work?*

An interesting issue for this kind of filtering is whether control and initiative lie primarily with the information consumer or producer. Filtering from broadcast data is a monitoring activity that can be controlled by the user or the system. In the example of the state of the assembly line, the user's interest has been encoded in the form of process control parameters, which are filtered by the system from the entire mass of instrument data. In the literature example, it is more likely that the criteria used to express interest are under the user's control. User profiles are relevant to filter designers here; given relevance feedback by users on the filter results, the profiles might be self-adaptive.

The question of who initiates a filtering request (user or system) is different from who is responsible for encoding the request. When the system initiates a filtering request, technology to support notification, such as triggers in active databases [Ris89], must be used.

If there are different possible criteria for selecting information to pass on to the user, then an important presentation issue for this kind of filter is how to indicate to the user which criteria caused the information to appear.

3.7 Multiple focuses with context

We have discussed the filtering goals of viewing trends and rationales and linking between multiple views. Each of these requires a user to look at different kinds of information at once in an exploratory way. In some cases, different pieces of information do not have a well-defined relationship, but are nevertheless held together by a user's common interest.

Example: *Which parts of the patient record should I look at before this patient's visit?*

Example: *What is the current course of treatment and what are the alternatives?*

To answer questions like this, multiple focuses are desirable. For the patient record example, a physician might want to see some general data on the patient along with the physician's own notes, treatments and test results. If the physician is a cardiologist, then he or she would probably not be interested in most of the ophthamologist's data but would be interested in drugs the patient is taking for any condition. However, a strict subset of information with no surrounding context is not desirable in this situation. By context, we mean closely related information that supports or helps to position the primary information. The context may be of interest to the physician after the primary information has been absorbed, or it may enhance the initial understanding of the primary information.

Generalized fisheye views are one obvious solution for filter designers to apply here. The access to context allows the physician to probe beyond the primary areas of focus so that even information with a low probability of relevance is still available, thus minimizing dangerous blind spots. The degree of interest function used in to construct a fisheye view might be static (as in the example above, based on professional specialty) or dynamic. The system might provide a fixed set of pre-constructed views or the user might control the creation of new views based on changing interests.

A simplified version of this filtering goal is when the user has a single topic of interest to explore.

Example: *Which of this patient's problems are active?*

Many single-topic questions can be answered by straightforward database queries. For other questions, such as the above example, the boundaries of the region of interest are ambiguous, both in specifying the filtering request and in presenting results. What does it mean for a problem to be active? There may not be clear rules to apply; the physician asking the question may want access to the gray area between problems that are clearly active and those that are clearly inactive. Again, fisheye views are a possible solution. Relevance feedback could also be used to offer a graduated view of elements in or nearly in the region.

4 Information Producer

We have shown that consumers have not one but a broad range of goals when they ask questions of large information spaces. These goals require consumers to obtain reduced views of the available information in order to bring out the useful subset, pattern, thread, rationale, or overview. From an infrastructure developer's point of view, different goal descriptions, information structures and presentation techniques are needed to bring out the relevant parts.

An understanding of the range of filtering goals can help the infrastructure developers give the consumer better tools for finding information. However, the consumer does not work with neutral information – the producer has already made many choices of what information is available to whom, in what form, and how it is structured. In this section, we will examine these choices.

Before analyzing how producers affect what information is findable and where an informed infrastructure can help support consumers' goals, let us clarify the differences between information producers and intermediaries.

First, an *information producer* is anyone who generates and records primary data. Thus, in medicine it includes the physicians, lab technicians, nurses, and authors of medical articles. In product development it includes the entire team – the usability engineer, simulation specialist, technician, electrical engineer, mechanical engineer, and so on. We also include the domain specialists who produce secondary sources, such as a survey article or videotape. However, we treat separately the information specialists whose role is to customize or transform information which someone else generates, such as the librarian who does custom searches or the news group moderator. We refer to the latter specialists as *information intermediaries.*

Besides the obvious choice of what information to record, producers must implicitly make complex choices of how to structure, disseminate, and present information. The choices are motivated by customary practice, technical practicalities, and partial understanding of consumers' goals. We will show how an explicit understanding of the consumers' goals can help the producer and the infrastructure developer better serve the consumer's needs.

4.1 Structure

The structure of information can be either formal or informal. By formal structure we mean data that is made up of named fields with values that can be meaningfully ordered or queried. Formal structure ranges from schemas for database records to the attributes under which an article is indexed, such as date, author, title or journal. The producer's responsibility is to choose appropriate values for such attributes.

In contrast to formal structure, complex information often has an informal structure that allows freedom of content within a conventional rhetorical framework. For example, an article describing a study may have a problem statement, a methodology section, results, discussion and conclusion. A physician's dictation about a patient encounter and problem history is prose, yet it follows a very stylized organization and content. The content includes conventional phrases such as "main complaint", "progression" and "differential diagnosis" that mark where answers can be found to overview, trend, and rationale questions.

Formal structure offers the strong advantage that the system can support efficient storage and search mechanisms using the elements of the structure. However, there is a cognitive load on the producer of information in designing this formal structure to support the right set of tasks. Just as some kinds of searches are made easy by adding structure, others are made very difficult. For example, patient records are stored in per-patient folders to support easy access to all of the information available on a particular patient; there is no structure overlaid on these records that allows searches for information across patients. Thus it took a pair of researchers four months to answer the question of who among a clinic's patients had a particular disorder.

One possible approach is to allow information to migrate from informal to formal structure, as indicated by common usage. Information Lens [MGT+87] takes this approach. Users can always use free-form text in messages, and they can also change the templates for messages if they discover patterns in how the free-form text is used.

A physician's conventional phrases are an example of a very useful informal structuring technique called beacons. Beacons are textual or pictorial guides that suggest the topics discussed nearby. For example, section headings help in skimming long articles. At a lower level, equations, tables, diagrams and images all serve as pictorial beacons. In engineers' lab notebooks, diagrams, figures, or underlined headings also serve as beacons that set up expectations for the content of a section. Confirmation of the importance of beacons comes from empirical studies of code comprehension [Bro83]. The CORE project [ELK+91] is exploring the value of diagrams of chemical structures or scheme drawings as a form of pictorial beacons, which chemists may use as they look through articles for descriptions of chemical transformations that are analogous to the ones they want to do.

Informal structure provides helpful filtering clues to readers. When information was primarily paper-based, sequential reading and skimming were the only activities that needed to be supported, hence structure and beacons were implicit in the text. However, implicitly-expressed informal structure is very difficult for a computational filter to use. Consider an online help query asking which of an application's commands change a certain screen parameter. Such a query may be impossible in a flat help document which does not express the implicit structure of commands and their descriptions in a computer-understandable way.

What does this imply for infrastructure developers? We believe that with appropriate infrastructure support, producers in domains such as product development or medicine would choose to annotate informally structured information in ways that would enhance its retrievability. In both of these domains, the producers often later become consumers; the physician tries to find things in the patient record to which he or she has contributed, and the engineer must understand how to change a design representation that he or she has helped create. Hence the producers are familiar with at least some of the eventual uses of the information.

What is needed to support producers in this annotation? First, domain analysis can identify the common filtering goals and the components of the conventional structure that can be marked or normalized for analysis. Second, user interfaces must make it possible for producers to easily add the appropriate beacons to the text, images, data sets, or video information. Third, the information infrastructure would have to allow both producers and consumers to declare or filter on these new data categories. Since producers are often too busy to systematically structure their information or may not realize all of the multiple uses for the information, a fourth strategy would be to use intermediaries or consumers themselves to classify and add structure. Medical transcriptionists can add structural markers to data, just as product design librarians can classify and annotate design documents for better retrieval. Currently, the closest model is that of MEDLINE cataloguing librarians who generate a rich set of attributes for classifying medical articles.

In summary, producers now use conventional structure and beacons to mark answers to different types of questions. Implicit structure is inaccessible to structure-based filters. However, flexible system support for both formal and informal structuring, along with an understanding of consumers' goals, can help information producers explicitly indicate structure.

4.2 Dissemination

Producers choose who should have access to their information, how they should get it, and how quickly. Let us explore the options for each of these choices and see how they interact with the choices of information structure and media.

First, the producer decides who should have access to the information. The choice may be inclusive (a fishing net) or exclusive (a barrier against the unwanted). An AIDS researcher selects a conference or journal at least partly on the basis of the publication's ability to reach the desired audience in a timely fashion. Other distribution choices are exclusive, based on the identities or roles of the potential recipients. These may be implemented by mechanisms such as file and database permissions, physical locks, or security ratings for physical documents. In making either type of access decision, the producer needs to understand the available information repositories and their attributes.

A second choice is whether to passively make the information available to interested searchers or to actively distribute it. This is actually a continuum, not a binary choice. Options range from a producer making a lab notebook entry or adding a page to the patient record to a very active mechanism such as a medical instrument alarm or a PA system. In between are conference presentations, electronic bulletin boards and telephone calls. The information infrastructure determines the choices available by providing connectivity, name services, distribution lists, database triggers or other alerts, or retry mechanisms for active distribution.

Distribution lists and specialized bulletin boards are examples of a common language by which consumers express their interest in a topic and producers make dissemination choices. Technology could help in allowing a richer language, in which consumers would indicate their level of interest in terms of their own goals (e.g. see all messages, track trends or critical events, or see an overview). Producers or intermediaries could then better describe their information to match the known goals.

The third producer choice we discuss is how quickly information should be spread. This interacts with the previous choice, since urgent, important information is likely to be distributed actively. Speed also interacts with notification. It does little good for a medical lab to quickly send test result to the requesting clinic if the physician is not told the result has arrived. The lab could phone a physician with a verbal report if it were known that a messenger would not deliver the X-ray or written report in time. Thus, to achieve the distribution speed goals, producers not only choose how to distribute information, but may transform it to allow a faster or cheaper channel.

Currently, a producer's dissemination choices often affect other decisions, such as the information content, structure, and the media the information is recorded in. Clearly, the content will be less filtered within a trusted context, such as a project group or a physician-patient encounter. Similarly, the structure and media are strongly affected. For example, electronic mail may give a fast, custom distribution, but one generally cannot include database records, links, or audio in messages. At a deeper information architecture perspective, these issues also interact with security, structure, and communication issues, including active databases, languages for user customization, and multimedia capabilities of different communications channels.

4.3 Media

The producer of information must choose the medium in which to express it. For both engineers and physicians the choices are becoming richer, with more complex tradeoffs. Tradeoffs are based on the information's accessibility, retrievability, utility and the cost to produce and store it in different media. This section examines these tradeoffs and their impact on usability.

The greater availability of audio, video, image and computer records makes it easier to store masses of information in a form closer to the original data, thus making it potentially easier to make comparisons or see patterns or trends. For example, in the last 15 years, audiotapes have been used effectively to teach infrequently-heard sounds relevant to diagnosis, such as whooping coughs, and videotapes are used to teach surgical procedures. Multimedia descriptions of orally- or visually-oriented information are more evocative than summary text descriptions. However, current patient records of gait problems, speech problems and operations are still transformed into text. Text provides an easily manipulable summary, and the producer who chooses to use a medium such as videotape faces difficult storage and indexing issues.

As it becomes easier to store and retrieve multimedia information, information producers can choose which parts of the data to store in each medium. They can add structure by providing markers to show critical incidents and annotations to give context. For example, a medical textbook might annotate X-rays or MRI scans to show pathology or anatomy. Structuring, classification and annotation of multimedia documents must currently be done interactively. Text media, in contrast, can be automatically classified using various techniques, e.g. latent semantic indexing [DDF+90].

Media transformation for retrievability is important, since it involves irreversible decisions of what is relevant and irrelevant. In time-oriented media such as video or audio, presentations of the data select the salient parts. As a specific example from engineering, the distribution of stress on a structure may be tested and visualized by polarized light transmitted through a stressed plastic model. An engineer must choose not only which of the test context and results to save in a project record, but how to show the results: photos in chosen perspectives, a videotape of the model reactions to changing stress, a diagram with isostress lines, or a textual description. Clearly, the choice will be based on the perceived purpose of the information and the cost of using different media. However, note that different choices make it either simple or impossible to answer questions about sequence or rationale or to make comparisons.

In short, transformations of multimedia information are difficult and involve loss of information. Producers must try to anticipate the filtering tasks that consumers want to do, such as obtaining overviews, identifying trends, making comparisons, or finding rationales for previous choices. Only by understanding user goals and using appropriate technology can producers make good choices of how to preserve information in both text and non-text media.

5 Conclusion

In this paper, we have offered insights into the roles played by the producer and consumer of information, a categorization of filtering tasks, compelling filtering examples from two domains, and a system design approach based on user needs analysis. We have developed a preliminary model of the factors needed to support filtering, including not only the producer and consumer roles, but also the intermediary, domain tasks and technology. We plan to refine this model and explore its applications in different domains.

6 References

[Bro83] Reuven Brooks. Towards a theory of the comprehension of computer programs. *International Journal of Man-Machine Studies*, 18, 1983.

[DDF+90] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41, June 1990.

[ELK+91] Dennis E. Egan, Michael E. Lesk, R. Daniel Ketchum, Carol C. Lochbaum, Joel R. Remde, Michael Littman, and Thomas K. Landauer. Creating and using an electronic library: Progress report on the CORE project. In *Proceedings of EDD'91 Bellcore/BCC Conference on Electronic Document Delivery*, 1991.

[Faf90] D. Fafchamps. Transcripts of ethnographic records collected in outpatient clinics of private practices and teaching institutions. Hewlett-Packard Laboratories, July 1990.

[Fur86] George W. Furnas. Generalized fisheye views. In *Proceedings of the Conference on Human Factors in Computing Systems, CHI'86*, 1986.

[FYT91] D. Fafchamps, C.Y. Young, and P.C. Tang. Modelling work practices: Input to the design and evaluation of a physician's workstation. In *Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care*, 1991.

[MGT+87] Thomas W. Malone, Kenneth R. Grant, Franklin A. Turbak, Stephen A. Borbst, and Michael D. Cohen. Intelligent information-sharing systems. *Communications of the ACM*, 30, May 1987.

[MRC91] Jock D. Mackinlay, George G. Robertson, and Stuart K. Card. The perspective wall: Detail and context smoothly integrated. In *Proceedings of the Conference on Human Factors in Computing Systems, CHI'91*, 1991.

[MSB90] John Alan McDonald, Werner Stuetzle, and Andreas Buja. Painting multiple views of complex objects. *ACM SIGPLAN Notices*, 25(10), 1990. (Proceedings OOPSLA/ECOOP '90).

[Ris89] Tore Risch. Monitoring database objects. In *Proceedings of the Fifteenth International Conference on Very Large Databases*, 1989.

[Tuf83] Edward R. Tufte. *The Visual Display of Quantitative Information.* Graphics Press, 1983.

[UY89] Kenneth Utting and Nicole Yankelovich. Context and orientation in hypermedia. *ACM Transactions on Information Systems*, 7(1), January 1989.

[Wil84] M.D. Williams. What makes RABBIT run? *International Journal of Man-Machine Studies*, 21, 1984.

[WSSC89] Sandra R. Wilson, Norma Starr-Schneidkraut, and Michael D. Cooper. Use of the critical incident technique to evaluate the impact of MEDLINE. Technical Report N01-LM-8-3529, American Institutes for Research, P.O. Box 1113, Palo Alto, CA 94302, September 1989.

[WTF+89] M.D. Williams, F.N. Tou, R. Fikes, A. Henderson, and T.W. Malone. RABBIT: Cognitive science in interface design. In *4th Annual Conference of the Cognitive Science Society (Ann Arbor MI)*, volume 8, 1989.