Protofoil: Storing and Finding the Information Worker's Paper Documents in an Electronic File Cabinet

R. Rao, S. K. Card, W. Johnson, L. Klotz and R. Trigg

UIR-R-1994-08

Protofoil: Storing and Finding the Information Worker's Paper Documents in an Electronic File Cabinet

Ramana Rao, Stuart K. Card, Walter Johnson, Leigh Klotz, and Randall H. Trigg

Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
<rao,card,trigg>@parc.xerox.com

ABSTRACT
Although the document imaging industry has taken off in the last few years, document image filing for the individual information worker is still not widespread or effective. In this paper, we focus on building an electronic filing system for paper documents that supports the ad hoc, multifarious work of information workers. Motivated by interviews with researchers and a survey of descriptive studies of paper document filing, we have focussed on minimizing or delaying costs of document filing and supporting a rich variety of methods for accessing and using stored documents. We have implemented a prototype system called Protofoil for storing, retrieving, and manipulating paper documents as electronic images that integrates many user interface—paper and workstation—and information retrieval technologies. Protofoil has been tested through use in our laboratory, and has been deployed in a field study at a lawyer's office.

KEYWORDS: Document Imaging, Paper User Interface, Information Retrieval, Filing of paper documents, Ad hoc information work.

INTRODUCTION
In the last few years with rapid improvements in compression, scanning, storage, and display technologies and associated decreases in costs, document imaging has taken off as an industry. However, as with other areas of paper-processing automation (e.g. accounting and billing), document imaging systems have primarily been directed at the enterprise or department as opposed to workgroups or individuals. Moreover, imaging and associated workflow applications have largely focused on automating highly-structured work in which many documents of a limited number of types are processed in the same manner (e.g. claims processing in insurance or check processing in banking). We have focussed on building an electronic filing system for paper documents that supports the rich ad hoc work of information workers. This kind of work differs in a number of ways: 1) documents vary widely in size, genre, content, purpose of filing, and context; 2) different document classes and even instances are processed and used in radically different ways; 3) up-front document filing costs can inhibit or even prohibit filing; 4) document use may occur long after document filing; 5) documents are usually gathered, filed, and used one or a few at a time.

CHI '94 Boston MA

The major research contribution of this work is in establishing requirements for supporting document imaging for ad hoc information work and in demonstrating how a variety of user interface and information retrieval techniques can be integrated and applied to these requirements. Two system goals have emerged as particularly important for supporting the rich variety of documents and uses in information work: first, minimizing filing costs or delaying them until they are justified by the user's needs; and second, supporting an extremely rich and varied process of document access and use.

We have developed a system called Protofoil that supports many of the paper filing and distribution needs of ad hoc information work. Protofoil includes functionality for capturing, retrieving, and browsing paper documents and invoking various document services including printing, mailing, faxing, and remote access. Several aspects of the Protofoil system design support the two principle design goals. Paper user interfaces [7] helps minimize and relocate costs of document scanning while still allowing a wide range of local and remote filing and access options. Furthermore, since a major part of filing-time cognitive cost is caused by the need to consider and plan for subsequent retrieval and use, a rich set of retrieval and browsing mechanisms can improve filing. Our previous work ([12, 13]) on an iterative multiple-method retrieval framework has been applied to the requirements of this application area. Protofoil has been used by a number of researchers and by a lawyer at a field site.

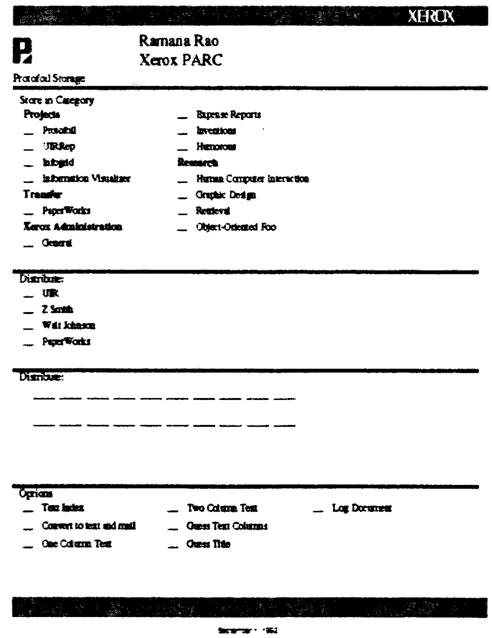
REQUIREMENTS
To help formulate requirements for building document imaging systems for ad hoc information work, we interviewed ten PARC researchers and a lawyer on their electronic and physical filing practices. A number of interview-based studies of the filing practices of information workers by others corroborate many of our observations [9, 4, 2, 3, 8].

All of the PARC researchers, importantly, had access to a previous research system [11] that allowed them to store paper documents in a shared database and to invoke a variety of services e.g. print, optical character recognition (OCR), fax. Most had used one aspect or another of this system, but not as extensively as expected. Two reasons in particular—the overhead of scanning documents into the system and an inability to find documents in the shared database—defined central thrusts for our system development activity. Below, we examine these two problems areas, document filing and document access, and outline our approaches.

Document Filing

Malone [9] found that piles of paper documents in information worker's offices are often a direct consequence of the mechanical difficulty of creating labeled file folders, binders, and so on. Like Malone, we too have observed the "piles problem:" paper piles up until some threshold of annoyance or inconvenience is reached and the worker faces the unpleasant task of purging and filing. Most current imaging systems substitute a different set of barriers to filing documents that limit use to high-volume, structured situations.

Our approach to minimizing scanning overhead is based on paper user interface technology [7]. Each user has one or more cover sheets (Figure 1) that can be placed in front of a document and put into a scanner's automatic document feeder. Various instructions or options can be selected by marking areas on the cover sheet and multiple documents can be separated by different cover sheets. Cover sheets offer several important advantages: 1) the user's time is decoupled from the scanning process, 2) remote usage (via fax) can be easily supported, 3) instructions and options can be marked anywhere, anytime, and 4) a display-based interface is avoided in walkup situations or on purely paper-based operations.


XEROX

Ramana Rao
Xerox PARC

Protofoil Storage

Store in Category

Projects
_ Protofoil
_ UIRRep
_ Infogrid
_ Information Visualizer

Transfer
_ PaperWorks

Xerox Administration
_ General

_ Expense Reports
_ Inventions
_ Humorous

Research
_ Human Computer Interaction
_ Graphic Design
_ Retrieval
_ Object-Oriented Foo

Distribute:
_ UIR
_ Z Smith
_ Walt Johnson
_ PaperWorks

Distribute:

Options
_ Text Index
_ Convert to text and mail
_ One Column Text
_ Two Column Text
_ Guess Text Columns
_ Guess Title
_ Log Document

Figure 1: A Protofoil cover sheet

A second barrier to filing observed by Malone is the cognitive difficulty of categorizing information in a way that supports effective retrieval. One of our subjects voiced a typical concern: "once something goes in there [a set of two 4-drawer file cabinets], it's gone." This typical subject thought very carefully about what and how to file to support future needs.

Retrieval technology offers important opportunities for addressing this problem, but it is important not to introduce new scantime barriers: for example, requiring the use of a workstation interface to assign titles or keywords at scan-time. The combination of paper user interfaces and a rich set of methods for browsing collections and finding documents improves the flexibility of when and how additional information is associated with a document. Cover sheets can have various indexing information (e.g. user, storage location) pre-encoded on them and markable regions of the cover sheet can allow for additional data or control information to be supplied by the user. A rich set of browsers and search techniques allows multiple, complementary approaches to locating desired information and interacting with the document corpus according to the needs of particular ad hoc tasks.

Document Access

A striking observation about how documents are stored in the office is the variety of factors that influence where they are and how people look for them. A common tendency in automation is to assume that semantic factors dominate, when in fact it is quite clear from many observational studies that they are often secondary or just one of many factors. Case [3] found that physical characteristics of both the office (spatial constraints and layout) and documents (physical form e.g. hardbacks, paperbacks, journals, articles) are given priority over topic in determining storage location in the office. Importantly, physical embodiment plays a crucial role in retrieval by providing a rich set of visual and physical cues. In addition, other contextual factors are important in document arrangement. Malone found that piles were often used to arrange materials as reminders of things to do and to make frequently used information readily accessible. Kwasnik [8] identifies other contextual factors including use or purpose (e.g. "to give a lecture ..."), circumstances (e.g. an ongoing project), and time (e.g. frequency of use or age of document).

The richness and creativity of the ways documents are arranged, sought, and used physically strongly suggests the importance of providing an equally rich and flexible environment for interacting with documents electronically. The Protofoil workstation interface is based on our previous work on user interface frameworks that integrate a multitude of information retrieval techniques and user interface design ideas [12, 13]. In this work, we instantiated particular versions of these various techniques suited to document imaging applications. Protofoil combines a category-based filing system with an associative database and a content-based retrieval and analysis engine. This allows finding a document using a number of techniques based on category (correlate of physical file folders), user-indicated attributes (title, keywords), system-derived attributes (scan-date, user actions), content (indexing and statistical analysis of OCR-ed text), and example (similar documents).

In addition to supporting multiple search methods, the Protofoil user interface supports multiple viewing methods for results and documents. Views can play a central role in helping refine retrieval as well as in various use tasks. A particularly important example is the use of reduced thumbnail images of pages. Such thumbnails recapture many of the visual cues of paper documents and can provide useful feedback to retrieval or even directly provide desired information.

SYSTEM ARCHITECTURE

A Protofoil workspace includes networked workstations, electronic storage, a digital scanner and printer, and a fax modem. The system topology is not rigidly defined: Figure 2 shows a

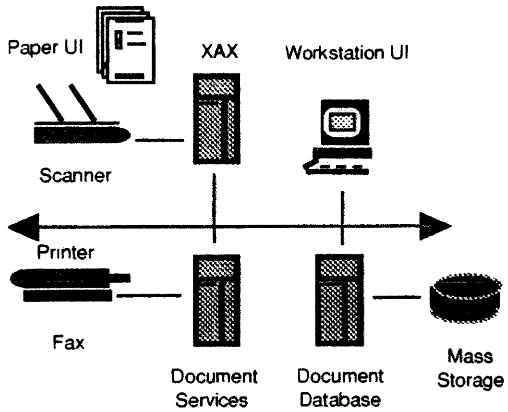


Figure 2: The Protofoil system consists of four major software components and various physical resources including a network, printer, scanner, electronic storage, and a computer workstation.

typical configuration in which each of various software components run on separate machines. There are four major software components:

- a *document database* for storing documents as page images along with other information needed to support search, retrieval, and browsing of the documents.
- *document services* for printing, emailing, faxing, and OCR.
- *scanstation user interface* for storing documents into the document database and for otherwise accessing documents or invoking services using paper user interfaces.
- *workstation user interface* for retrieving and browsing documents and auxiliary information and invoking services.

The first two components provide backend document functionality and the other two provide frontends (i.e. user interfaces). The frontend/backend separation enables us to provide user interfaces in both the paper and electronic domains as well as multiple interfaces in each of these domains tuned for different uses. We describe the document database and services here and particular incarnations of frontend components in following sections.

The document database provides a repository for storing documents and associated information. Each user has a separate area in the document database, but a user can choose to allow access to particular documents or sets of documents to other users. The major operations supported by the database include storage, deletion, search of various kinds, and retrieval of documents.

Documents are stored by providing a set of page images captured at a given resolution (typically 200dpi for fax and 300 or 400 dpi for scanned) and a description record. A document's description record consists of various required and optional fields. Required fields include User, Category, Scan-Date, Scan-Location, and Number of Pages; and optional fields include Categories, Title, and Saved Regions. Optionally, optical character recognition (OCR) is used to extract text from the page images and the text is added to one or more user or group indices. These indices are maintained and accessed through a full-text engine called TDB [5] that supports word stemming and a variety of text-based search techniques.

In addition to the page images, the description record, and recognized text, the database generates and stores reduced page images at a number of different sizes. These reductions are used to support various interactions in the paper and screen user interfaces (e.g. thumbnails and screen-reading). Reductions are generated using standard raster scaling algorithms as well as special algorithms suited to various document types or properties (e.g. textual documents can be scaled in a way that enhances their recognizability).

Several document services are supported including printing, faxing, and emailing documents as well as "publishing" a document into shared areas of the document database. Convenient access to these various services on electronically stored document is an advantage over a physical filing system.

PAPER USER INTERFACE

Paper user interface technology is used to store documents into the database as well as to retrieve documents and to invoke document services. To store a document, the user uses a form as a cover sheet on a document, possibly filling in fields. For example, the form in Figure 1 is used by one of the authors to store documents. Many description fields are generated automatically by the system, reducing user overhead during filing. The user can select categories for a document at filing time or mark special regions that are clipped and saved with the document. Constrainted handprint recognition is used in some fields and various heuristic extraction from the document itself has been explored (e.g. title extraction).

A major advantage of paper forms is that they can be utilized in many familiar ways (e.g. copying and reusing partially completed form). In addition, the user can have one or more forms for storing documents or invoking other services. New forms are created by editing structured descriptions or by using a WYSIWYG editor. They are then registered with the system. The multipurpose form, shown in Figure 1, supports storage into predefined categories; redistribution via fax or email to predefined individuals and groups or via fax to a handprinted phone number; and other options including recognition options, text indexing, logging, title guessing.

The central system component for processing paper user interfaces is the XAX paper server [7]. XAX processes page images that are placed in its spool directory by a scanner or a fax modem. XAX recognizes paper user interface forms and interprets data encoded as *glyphs* to identify the form. XAX then processes the user-markable regions on the form and invokes appropriate actions on subsequent page images along with user-specified arguments and options.

WORKSTATION USER INTERFACE

The Protofoil workstation user interface is based on our previous work on information access user interface frameworks [12, 13]. The interface, shown in Plates 1 and 2, uses a tiled pane layout with three columns: the thin left column containing tools for invoking operations; the middle column for specifying retrieval requests and browsing search results; and the right column for browsing documents.

Document retrieval is based on an iterative loop in which four user actions are variously interleaved: selecting a scope (initially the entire file cabinet); specifying a query; browsing results; and browsing and using documents. The user starts a retrieval by filling in a query form and clicking on the query tool. A query can use a variety of boolean and textual search techniques, singly or in combination. The multiple methods offer a range of strategies for locating particular or task-relevant documents. Five search methods were incorporated:

- Category Search matches on one or more categories. Categories can be assigned to a document at scantime using a checkbox on the cover sheet or by using the screen interface while browsing a document.
- Associative Search matches on document description fields including title or keywords. Description fields can be added or changed using the document browser.
- Similarity Search matches documents based on query text. To specify this search method, the user types words into the "Text" field.
- Similarity Search based on specific documents. After useful documents have been found, similar ones can be found by putting them in the "Similar To" pane.
- Proximity Search matches documents in which a bag or sequence of words occur within a given number of words. The user types words into the "Text" field along with an interger representing the size of the proximity distance.

In addition to multiple search methods, the Protofoil interface provides multiple ways to browse search results and documents (buttons in title bars switch among different views of pane contents). The interface attempts to automatically select the most appropriate view based on search technique. Five views of search results are currently supported:

- Thumbnails—an array of thumbnails (this visualization is shown in the "Similar To" pane of Color Plate 1). Thumbnails, to some extent, supports the kind of rapid visual recognition tasks often used in leafing through paper documents.
- Description—a list showing attributes of each document similar to the list view used in the Macintosh Finder (and other file managers). This view is automatically selected on Similarity Search and is augmented with scores and a list of matching words.
- Category Groups—a grouping of documents according to their category, automatically selected on category search. The categories can be expanded in this view to show matching documents in each category. Since categories are user-defined, this view organizes search results in a user-specific way.
- Clusters—a list of automatically-generated clusters of documents shown in Color Plate 1 (uses technique of [6]). Each cluster item has a reference number, the number of documents in the cluster, the central terms of the cluster, and a few central documents for the cluster. Clusters can be expanded to show included documents. Clustering provides an alternative to manual categorization for grouping documents.
- Snippets—a list of documents returned by a proximity search along with matching locations. This view is based on the snippet design [10] for showing keywords in context. In Color Plate 2, a proximity search on "hostile takeover" has been performed. Each snippet highlights and centers the nearest additional "useful" word in that region, so that the user can quickly assess relevance.

Result Sets or Document Groups (category groups or cluster groups) can be used as the scope of a further cycle of query. In Color Plate 1, the 6th cluster from a previous result in cluster view has been selected as the scope (more groups can be added), and an empty search, which matches all documents, was performed on this scope into the cluster view, which scatters the 189 documents into a new set of clusters (the documents in the "Similar To" were added after this last step). This particular interaction loop is called scatter-gather in [6] and is a special case of the more general iterative interaction cycle of setting scope, searching, and viewing results.

Once particular documents are located they can be placed in a "Holding Area" for later use, dropped on the "Document Tools" like "Print" or "Fax," or browsed using the panes on the right side. The "Document" pane supports four different views of a document: a reduction of the scanned pages appropriate for full-page screen viewing (shown in Color Plate 1), two grid of pages views at two levels of reduction (smaller size shown in Color Plate 2), and OCR-ed text view of the document. Clicking on a snippet automatically selects the text view, and locates and highlights the snippet within the document.

LABORATORY USE

Protofoil has been used extensively by three users and to a lesser extent by eight other users in our lab. Usage has centered around managing collections of disparate clips from trade magazines, or short documents involved in administrative activities like resumés, purchase orders, registration forms, etc. We were not able to provide scanners robust enough to handle the typically longer two-sided documents common in our research environment.

The top four Protofoil users stored over 500 documents, totaling over 1500 pages. These users, all of who used scanners previously, confirmed that document entry was simplified by the elimination of a screen user interface for scanning. Use of the forms to provide up-front categorization of documents and options for document processing varied across users. One pattern was to reuse a small set of pre-filled forms, selecting from these forms depending on the document and task. In this pattern, only a few common categories are selected up-front, and other categorization decisions are delayed. This usage pattern was supported by refining Protofoil to provide an in-basket category that automatically found all uncategorized documents so that they could be categorized as desired. OCR-text-based search methods provided supplementary mechanisms for narrowing the search or for locating documents directly. Though OCRed text doesn't currently play a major role in commercial document imaging system, its use in our laboratory setting indicates that it will be very important in ad hoc work applications.[1]

Others used forms in more varied ways, particularly to invoke other services on the document at the same time as storing the document. These users changed their forms to better suit their needs as they understood the systems capabilities and limitations (e.g. adding OCR options to improve recognition quality). These users also tended to add and delete categories in new revisions of forms. Another common occurrence was to split storage into multiple forms, for example to support different task situations (e.g. in the office vs. while travelling).

Besides scanner-related barriers, the two major impediments to broader use of Protofoil for filing are the lack of adequate storage infrastructure and the lack of a guarantee that such files will persist, or more accurately, remain accessible for as long as paper files would. The first problem can currently be solved at reasonable cost. The second raises more complex social and technological issues that require further attention.

[1] OCR need not be perfect to play a useful role in retrieval: particularly, because search words (keywords) are often used multiple times in relevant documents.

FIELD STUDY

Many of the document access capabilities of Protofoil were assessed in a field study at a lawyer's office in collaboration with the Work Oriented Design project.[2] The project involves studies of work practice at specific sites in conjunction with development of site-customized prototypes. In this particular field study, the Protofoil workstation interface was demonstrated to and used by an attorney, referred to here as Mitch, at a law firm using his own document collection.

Mitch maintains a corpus of a few thousand "model" documents (4 lateral drawers). These have been gathered over the course of years to support the process of document creation. Besides supporting Mitch's own practice, they have become a shared resource in the firm. The files serve a number of purposes. Some, like journal articles and copies of tax legislation are useful for the information they provide. Others capture the form of particular kinds of document, say a buy-sell agreement. Still others contain actual language that can be reused in a new document. The result is an extremely heterogenous "form file" of documents. Genres include memos and letters, faxes, online printouts of Lexis online records, stock certificates, contracts. Some are pure text while others include graphics. Some are in color; some are handwritten or have handwritten marginaliz; a few even include post-its. The corpus is constantly changing as new documents are added and (occasionally) old ones thrown away.

The field assessment of Protofoil included a pilot demonstration using a small collection of Mitch's documents, refinement of the interface, demonstration on a larger set of documents, and finally leaving the prototype for use in the course of actual work needs. The pilot demonstration included 36 documents (459 pages) from five folders, with names like "Beta site agreements," "Debt financing," and "Delaware Corporate Law." This was far too few documents to adequately demonstrate clustering or for that matter search, but enough to give Mitch a sense of the interface and to initiate discussions. We modified Protofoil in response to what we learned from these discussions.

The final corpus included one fourth of Mitch's files: 1 lateral drawer with confidential files removed and high use files from other drawers added for a total of 6794 pages of 862 documents taken from 94 folders. We left a workstation running the improved Protofoil in Mitch's office alongside the physical file cabinets with a video camera set up for recording encounters. Mitch used the prototype on his own for approximately two weeks in a number of recorded play sessions as well as real encounters (there were some unrecorded encounters). Some of these encounters also involved Mitch demonstrating the prototype to colleagues or helping them use the prototype in actual retrieval tasks. Initial review of the videotape has convinced us of the tremendous value of leaving behind the system running on an actual corpus used in "real" tasks including some involving interactions with coworkers.

While two demonstration and subsequent usage by Mitch do not provide a comprehensive assessement of the Protofoil interface, they have produced a number of validations and insights. At the most basic level, Mitch and others were able to use the prototype based on only two demonstrations (the small amount of documentation was hardly touched) and in most cases, were able to perform their search tasks effectively. Conversations with Mitch and the videotaped interactions confirm that Mitch believes that our interface elements would be useful in an electronic version of his form file, if properly integrated with his current electronic world. Though a more detailed analysis of the field study is forthcoming, a number of specific areas are worth raising here:

Image Views. We introduced new page image reductions based on the pilot demonstration. Originally, we used two sizes: one in the Thumbnail view for browsing results and the other in a screen-sized page view for browsing documents (both shown in Color Plate 1). Though Mitch initially stated that he couldn't see much in the thumbnails, he and a colleague went on to use them extensively in identifying aspects of their documents. Besides identifying the types of many documents, he was able to recognize particular features of some documents (e.g. the shape of company letterheads). To provide further support for document and page recognition, intermediate page image sizes were introduced to allow Mitch to browse a document at a level appropriate to his purposes (e.g. locating a particular page vs. reading the text). Mitch's use, as well as our laboratory study, indicates that direct physical rendering of document pages at various reductions can retain some of the visual advantages of paper documents.

Stop List Control. Stop words, words that are ignored during indexing and search, can dramatically impact retrieval effectiveness. In particular, the pilot demonstration quickly revealed that a standard English stop word list was inadequate for Mitch's setting. Words like "corporation" or "agreement" are no more useful for distinguishing documents in Mitch's corpus than words like "and" and "the." As the consequence, similarity and clustering results were significantly degraded. In particular, these words often appearred in the cluster or ranking displays, although for lawyers in this firm, they didn't help discriminate. In our design, we added a stop word list editor together with a button to re-index the text of the corpus (a 20 minute operation in his case). The more general results are: 1) an important requirement for retrieval and clustering techniques is greater flexibility in the use of stop lists (e.g. supporting operation-time stop lists in addition to indexing time stop lists) and 2) more attention should be given to generating useful results or summaries (e.g. by sensing non-discriminating words during indexing).

Example-based Searches. Relevance Feedback, using example documents to seed subsequent cycles of search, has become a widely used technique in the text retrieval community. Mitch's use of this technique raised important questions about its effectiveness in an electronic file cabinet application. In particular, Mitch might have been attempting to locate similar documents along dimensions other than the one supported by textual similarity search (e.g. image- or genre-based similarity). This raises the question of how best to support other forms of example-based searches and how best to design and represent example-based searches (including text-based techniques).

Search/View Interaction. Protofoil supports multiple search methods as well as multiple result visualizations. The field study underscored how strongly these two sets of tools interact. For example, we refined the manner in which views are automatically selected based on search method. In addition, the category view was developed specifically for category search. These design improvements all point to the deeper insight that visualization

[2] Work Oriented Design [1] is a project conducted by Jeanette Blomberg, Lucy Suchman, and Randall Trigg aimed at the integration of work practice analysis and cooperative design of applications with end users.

is complementary to search. For example, Mitch wanted to perform a similarity search restricted to a particular category. One way to do this is by selecting a category as the scope and then performing the similarity search. However invoking the category view on the results of an unrestricted similarity search returns the desired result set along with other possibly useful information (e.g. similar documents in other categories). Further work is necessary to understand the implications of interleaving search and visualization.

DISCUSSION

The burgeoning commercial document imaging market includes products from IBM, Wang, Filenet, ViewStar, Plexus, and others with a revenue of 1.9 billion dollars in 1992. Almost all of the dollar volume derives from high-end sales to paper-intensive vertical industries in support of repetitious routing and processing of documents. Many system issues of integrating scanning, storage, OCR, routing, retrieval, and browsing have been addressed in these high-end structured-work imaging systems. The Protofoil system shifts focus from automating repetitive processing of limited document types to the more varied work performed by researchers and other information workers.

The value of imaging in information work at the individual level as well as the group level has been recognized by others. The Canofile 250 product from Canon provides a stand-alone document filing system with optical storage, scanning and printing, and a browsing station, but isn't integrated with electronic environments. Image-enabled Notes from Lotus and Kodak allows image documents to be shared and reused in an organizational corporate document store. A number of others, most notably Keyfile Corp and Westbrook Technologies, have integrated scanned documents into electronic document management systems. Though these systems have gone a long way toward minimizing document capture costs, they still require up-front effort to file documents. Most of these systems include simple search capability on OCRed text, but they have not applied other image processing or recognition technology. They have also not provided as rich a process of interactive, iterative retrieval and browsing of results or documents.

Protofoil focuses on a number of specific problems that must be addressed in the pursuit of an electronic filing system for paper documents for information work. First, paper user interface technology provides a powerful means for easing the cognitive and mechanical overheads of the filing bottleneck. Second, by providing a host of search methods and document services in the electronic world, Protofoil offers greater flexibility during filing as well as a richer set of strategies for finding and using documents.

Many of the files in a personal file store can be eliminated with increasing electronic availability of various materials. However, in our view, there will be paper documents that are shared, delivered, carried, browsed, and digested for a long time to come. As long as this is the case, smoothing the process of allowing paper documents to enter or return to the electronic world and supporting facile retrieval and use of these documents despite their physical origin are important pursuits.

ACKNOWLEDGEMENTS

We thank the interviewees and Protofoil users at PARC, and our field user, "Mitch." Doug Cutting, Jan Pedersen, and David Karger provided us a very powerful search engine and clustering techniques. Larry Masinter and Steve Putz provided early document imaging infrastructure at PARC.

REFERENCES

1. J. Blomberg, L. Suchman, and R. Trigg. Reflections on the work-oriented design project in three voices. In *Social Science, Technical Systems, and Cooperative Work*. 1994. Forthcoming.

2. D. O. Case. Collection and organization of written information by social scientists and humanists: a review and exploratory study. *Journal of Information Science*, 12:97–104, 1986.

3. D. O. Case. Conceptual organization and retrieval of text by historians: The role of memory and metaphor. *Journal of the American Society for Information Sciences*, 42(9):657–668, 1991.

4. I. Cole. Human aspects of office filing: Implications for the electronic office. In *Proceedings of the Human Factors Society 26th Annual Meeting*, pages 59–62, 1982.

5. D. Cutting, J. Pedersen, and P.-K. Halvorsen. An object-oriented architecture for text retrieval. In *Conference Proceedings of RIAO'91, Intelligent Text and Image Handling, Barcelona, Spain*, pages 285–298, April 1991.

6. D. R. Cutting, D. R. Karger, J. O. Pedersen, and J. W. Tukey. Scatter/gather: A cluster-based approach to browsing large document collections. In *Proceedings of SIGIR'92*, June 1992.

7. W. Johnson, S. K. Card, H. Jellinek, L. Klotz, and R. Rao. Bridging the paper and electronic worlds: The paper user interface. In *Proceedings of INTERCHI, 1993*, pages 507–512. ACM, April 1993.

8. B. H. Kwasnik. *The Influence of Context on Classifactory Behavior*. PhD thesis, Rutgers University, 1989.

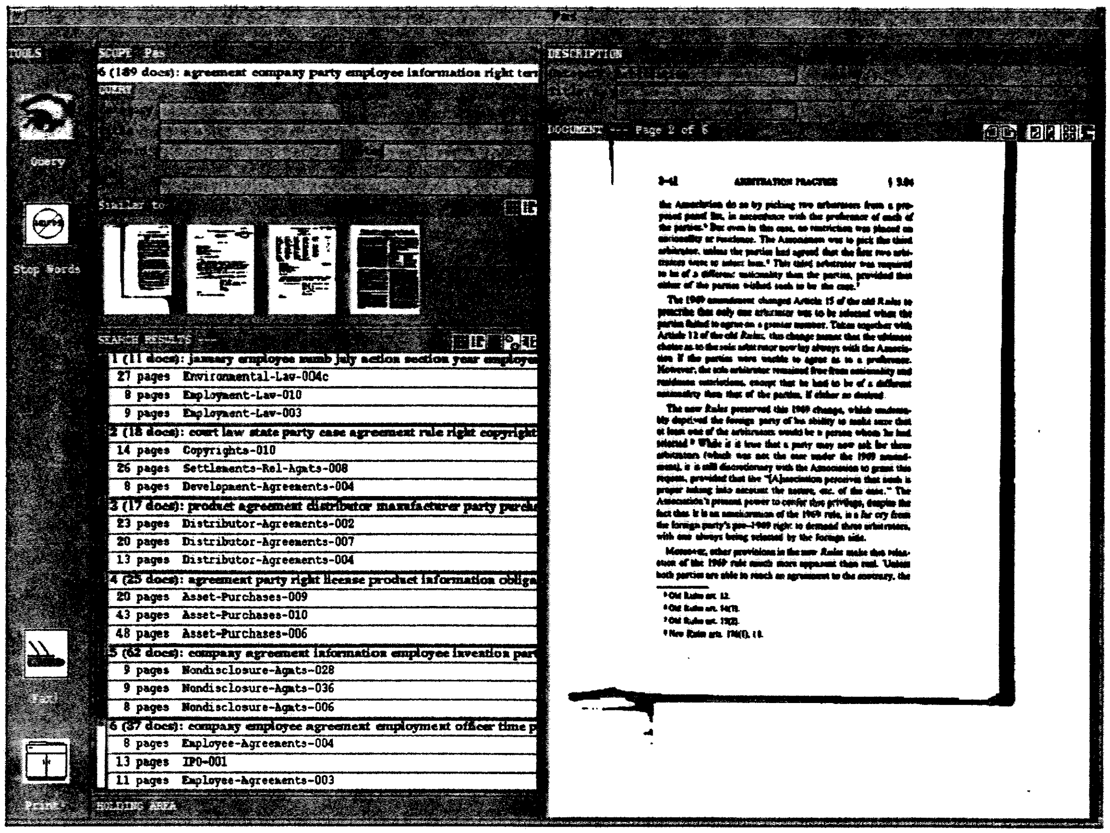
9. T. W. Malone. How do people organize their desks? implications for the design of office information systems. *ACM Transactions on Office Information Systems*, 1(1):99–112, 1983.

10. J. O. Pedersen, D. R. Cutting, and J. W. Tukey. Snippet search: a single phrase approach to text access. In *Proceedings of the 1991 Joint Statistical Meetings*. American Statistical Association, 1991.

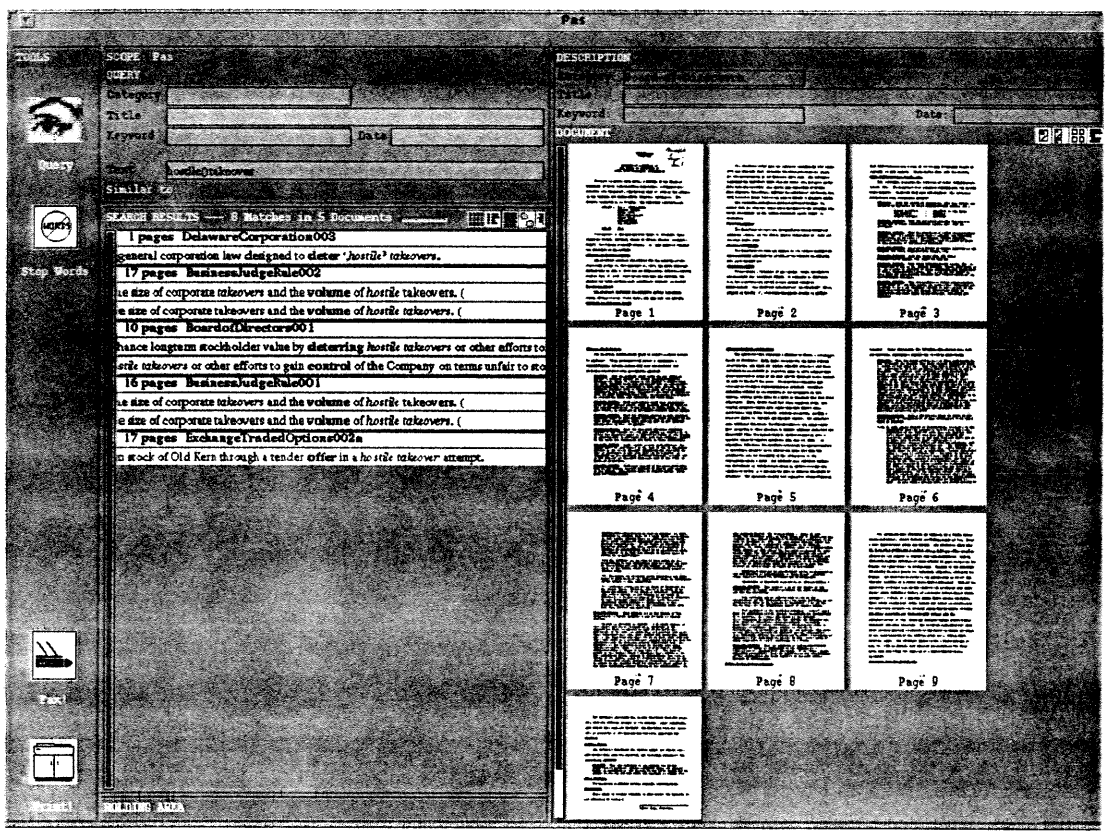
11. S. Putz. Design and implementation of the system 33 document service. ISTL Tech Report P93-00112, Xerox PARC, ISTL, 1993.

12. R. Rao, S. K. Card, H. Jellinek, J. D. Mackinlay, and G. G. Robertson. The information grid: A framework for building information retrieval and retrieval-centered applications. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM Press, Nov 1992.

13. R. Rao, D. M. Russell, and J. D. Mackinlay. System components for embedded information retrieval from multiple disparate information sources. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM Press, Nov 1993.



Rao, Card, Johnson, Klotz, Trigg, Color Plate 1



Rao, Card, Johnson, Klotz, Trigg, Color Plate 2