

Seamless Capture and Discovery for Corporate Memory

David M. Hilbert, Daniel Billsus, Laurent Denoue

FX Palo Alto Laboratory, Inc.
3400 Hillview Ave., Bldg. 4
Palo Alto, CA 94304 USA
+1 650 813 7765

{hilbert,billsus,denoue}@fxpal.com

ABSTRACT

In a landmark article, over a half century ago, Vannevar Bush envisioned a “Memory Extender” device he dubbed the “memex” [7]. Bush’s ideas anticipated and inspired numerous breakthroughs, including hypertext, the Internet, the World Wide Web, and Wikipedia. However, despite these triumphs, the memex has still not lived up to its potential in corporate settings. One reason is that corporate users often don’t have sufficient time or incentives to contribute to a corporate memory or to explore others’ contributions. At FXPAL, we are investigating ways to automatically create and retrieve useful corporate memories without any added burden on anyone. In this paper we discuss how ProjectorBox—a smart appliance for automatic presentation capture—and PAL Bar—a system for proactively retrieving contextually relevant corporate memories—have enabled us to integrate content from a variety of sources to create a cohesive multimedia corporate memory for our organization.

Categories and Subject Descriptors

H.3 [Information Systems] Information Storage and Retrieval; H.5 [Information Systems] Information interfaces and presentation (e.g., HCI); H.5.1 [Information Interfaces and Presentation] Multimedia Information Systems; H.5.3 [Information Interfaces and Presentation] Group and Organization Interfaces; C.2.4 [Computer-Communication Networks] Distributed Systems

General Terms

Algorithms, Human Factors

Keywords

Corporate memory, organizational memory, automatic presentation capture, proactive contextual retrieval

1. INTRODUCTION

Corporate memory has been a hot topic in management and computer science since at least the early 1990’s [25]. Despite great interest in technologies to help organizations leverage past experience, expertise, and other collective resources, in practice, corporate memories often fall short of expectations. There are numerous challenges in developing effective corporate memories, including problems in asset capture, representation, retrieval, and reuse. For instance, most people will not do extra work to contribute to a corporate memory without proper incentives,

particularly when the benefits accrue to someone else at some later time. Fortunately, many organizations already produce useful information that can be hosted on corporate intranets and tapped for corporate memory purposes. However, given the variety of available information and users in even a moderately sized organization, it becomes extremely difficult for employees to stay abreast of all the potentially useful information that may be helpful at any given time. This problem is exacerbated by the fact that sources of corporate information are often heterogeneous and distributed. Users must not only realize that there may be useful corporate information relevant to their current task, but they must know where to look, and possibly consult multiple sources.

Problems of resource discovery and information overload have plagued us for some time on the World Wide Web. With so many online resources available to us, how can we stay abreast of all the potentially useful information without spending all our time searching and filtering? In a corporate setting, these problems have palpable costs. When we are unaware of relevant information and human resources, the quality, efficiency, and satisfaction of our work suffers. When separate corporate units are unaware of one another’s activities and expertise, they unnecessarily duplicate work. As a result, systems that enhance resource discovery while limiting information overload have received much attention in both academic [6][20] and commercial contexts [4][24].

The goal of FXPAL’s research on Corporate Memory is to produce technology for more effective knowledge capture and utilization. We are investigating techniques that allow employees to contribute and access information seamlessly, i.e. without significant usage overhead. We believe that this can be accomplished by making both information capture and retrieval part of established work practices. For example, giving a presentation in a conference room, bookmarking a web page or sending email to a colleague are activities that result in corporate content that can be captured automatically. Similarly, activities such as accessing a web page or an email message can be interpreted as an implicit expression of an information need, which allows us to retrieve potentially relevant information proactively, i.e. without explicit user requests [5].

In the following section we provide a brief historical account of several systems created and used in our lab to capture, share, and communicate information within the organization. We then identify some problems that limited their usefulness in providing an effective corporate memory. The rest of the paper focuses on two new research systems that have recently improved our ability to index, retrieve, and integrate content from these various sources to create a cohesive multimedia corporate memory for our organization.

2. CORPORATE MEMORY AT FXPAL

Since its inception 10 years ago, FXPAL, a 50-employee research lab for Fuji Xerox in Palo Alto, has experimented with a variety of technologies for capturing, sharing, and communicating information within organizations. In this section we briefly describe four technologies that have subsequently become part of our internally deployed corporate memory system.

2.1 PALWeb

Like many other organizations, FXPAL has an internal web site that provides unified access to a range of corporate resources, including an employee directory, group calendars, a meeting room scheduler, links to corporate policy, forms, and procedures, and a document repository for our “Products”. Products include Technical Reports, Technical Memos, Invention Proposals, Software Releases, and Publications. PALWeb provides a useful, database-driven repository and portal for official corporate resources.

2.2 Video Guestbook

The Video Guestbook [22] is a self-service kiosk designed to record information about people who visit our organization. The kiosk consists of a conversational character running on a touch screen display (to greet and instruct visitors), a business card scanner (to capture visitors’ contact information), and a webcam (to capture visitor faces and pronunciation of their names). All captured content is stored in a shared contact database that is accessible to users via a web interface, and to other systems via a web service API.

2.3 Plasma Poster

The Plasma Poster Network [9] is a collection of large screen, digital poster boards used to display community-contributed content in public places throughout an organization. It was inspired by the use of physical bulletin boards and designed to stimulate unplanned social interactions around digital content, possibly leading to opportunities for discovery of overlapping interests across work groups. Employees can simply email web pages or digital photos to the Plasma Poster Network to publicly post content.

2.4 UbiSight

UbiSight [16] is an automatic video capture and composition system. It dynamically steers multiple pan/tilt/zoom (PTZ) cameras and selects the best close-up view for passive viewers. Its purpose is to maximize captured video information with limited cameras. In essence, UbiSight acts as a “virtual camera man”: it uses motion and audio tracking to automatically find regions of interest. At FXPAL, the system records nearly all meetings in our main conference room: since very little manual intervention is required (currently we only have to start and stop the system), the system generates high-quality video of presentations with limited user effort.

2.5 Enabling Corporate Memory

The above systems created and managed content that could be useful for a corporate memory. However, the content produced by these systems was not easy for employees to access. The video recordings were not indexed to allow segments of interest to be retrieved based on content. Also, the onus fell on users to know where to look for content and to make themselves aware of what was new and relevant to their current tasks. Finally, these systems produced “information silos”, meaning users not only needed to

manually search for content, but needed to consult multiple sources to gain a complete picture. In other words, we lacked effective support for promoting resource discovery and managing information overload. As a result, we did not utilize our resources effectively: FXPAL’s employees had to work pretty hard to make use of this content.

In the following sections we describe two research prototypes that have recently allowed us to overcome these problems. ProjectorBox, an autonomous presentation capture appliance, now indexes virtually every presentation and meeting in our corporate conference room, allowing video segments to be retrieved based on content. And PAL Bar, our proactive retrieval system, brings together content from these heterogeneous sources into a unified multimedia corporate memory.

3. PROJECTORBOX

Presentations are ubiquitous in education, business, and government. But presentation archives are rare due to the cost of purchasing, setting-up, and using current recording technology. As a result, useful information passes through projectors all the time and is lost. If we could create useful archives cheaply and easily—*without any added burden on anyone*—the benefits would be far reaching.

Current approaches assume that presenters, facility operators, or audience members will adjust their practices (sometimes slightly, sometimes significantly) to garner the benefits of automatic presentation capture. This includes: Georgia Tech’s Classroom 2000 [1], Cornell’s LectureBrowser [18], Bellcore’s STREAMS [10], PARC’s meeting capture tools [17], FXPAL’s early video capture work [8], Microsoft Research’s automatic camera control work [15], AutoAuditorium [3], Sonic Foundry’s MediaSite [21], and AnyStream’s Apreso Classroom [2]. In our experience, even the most modest assumptions—e.g., that presenters will use specific software (e.g., Quindi [19]), upload their files, or start and stop recordings—are unrealistic. ProjectorBox is our attempt to create a presentation capture solution that users can simply “plug-in and forget”.

ProjectorBox captures, indexes, and archives presentation slides and audio without any user input. Like MediaSite [21] and Apreso Classroom [2], it is an RGB-based appliance that intercepts video information as it is sent from presentation devices, such as a presenter’s laptop, to display devices, such as a projector. This means it can capture content from any presentation device—including a non-preconfigured guest presenter’s laptop—running any presentation software. However, it goes beyond existing approaches by automatically creating easily searchable and skimmable archives without anyone having to start and stop (or schedule) recordings.

Despite their advantages, RGB-based solutions introduce considerable challenges. In order to automatically produce high-quality archives suitable for browsing, retrieval, and reuse, an RGB-based solution must infer high-level information from the video signal. For instance, the system must: (a) distinguish between presentation and non-presentation content, to filter unwanted content and start and stop recordings automatically, (b) detect boundaries between contiguous presentations to structure content for browsing and retrieval, and (c) extract text to support content-based retrieval. We now briefly discuss how we address these challenges, and provide implementation and user interface details in subsequent sections.

3.1.1 Detection of Presentation Content

Our first challenge was to automatically separate presentation content from non-presentation content and to free presenters from having to remember to start and stop recordings themselves. Researchers have noted the importance of not distracting instructors with new recording technologies, particularly at the beginning and end of classes when students often ask questions [1]. And our own experience [8] has demonstrated that if people must remember to start and stop recordings, then most presentations will simply not be recorded. In terms of RGB capture, this meant we needed to robustly classify screen activity as either “associated with a presentation” or as desktop activity “not associated with a presentation”.

For example, when recording a presentation, we would like to exclude images of the desktop and the presenter selecting the presentation file and starting the presentation software. We address this problem using an automated image classification approach. As a first approximation to identifying presentation content, we focused on identifying slides. We apply commercially available optical character recognition (OCR) to extract the text from images, and use the text bounding box heights as the main feature for classification. The intuition is that most presentation slides use larger fonts than typically appear on a computer desktop or in other applications. As a result, text bounding box heights are a useful cue for detecting slide images. The resulting classifier can reliably distinguish between presentation and non-presentation content [13].

3.1.2 Presentation Segmentation

Because ProjectorBox may run continuously in a room used by multiple people for multiple presentations, we also needed to automatically detect presentation boundaries to allow them to be browsed and retrieved as cohesive units. After some experimentation, we adopted a simple time-based approach in which presentation boundaries are defined whenever the elapsed time between two successive slide images exceeds a user-definable threshold (e.g., 20 minutes in our corporate conference room). This approach emphasizes precision in segmentation over recall. That is, adjacent presentations may be erroneously grouped together, but a single presentation is rarely erroneously split into multiple presentations. We make it easy for users to manually adjust presentation boundaries, and are actively experimenting with other approaches to improve segmentation.

3.1.3 Presentation Retrieval

Finally, students want to retrieve lectures based on content and access captured media non-linearly, as opposed to having to play through it sequentially [1]. We also experienced similar requirements in our own corporate conference room [8]. As a result, we use commercially available optical character recognition (OCR) to extract text from slide images and create a full-text index so users can retrieve slides based on content. And our slide skimming and playback interfaces, described below, allow users to easily skim and skip around presentation content non-linearly.

3.2 Implementation

ProjectorBox is a PC-based appliance equipped with a high-resolution VGA capture card and microphone. The capture card [11] can record VGA signals from any computer at any resolution up to 1600x1200. The system can record audio streams using any Windows-compatible audio device. We have installed our prototype in two different small-form-factor PC cases, which are

easy to deploy in conference rooms and classrooms and can be integrated with existing presentation podiums (Figure 1).

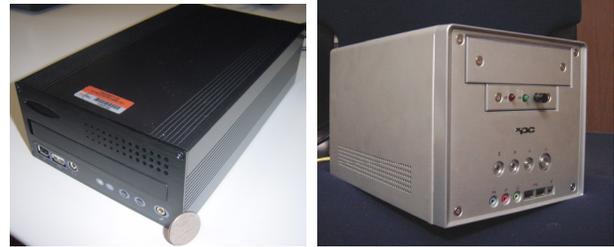


Figure 1. Two ProjectorBox prototypes.

ProjectorBox consists of two main software components: a capture component and a server (Figure 2). The capture component periodically transmits data to the server for further analysis and storage. Since presenters often flip back and forth through slides non-linearly, the capture component sends an image to the server only after it has been shown for several seconds. Based on usage in our corporate conference room, a 4 second interval produces reasonable results: slides that were not meant to be presented or discussed are omitted, while important presentation content is preserved. For each captured image, a corresponding audio clip is recorded from an external microphone and stored as an MP3 file. In our conference room, we use a ceiling microphone positioned above the center of the room. In other deployments, we have used centrally positioned table microphones or standard low-cost PC microphones.

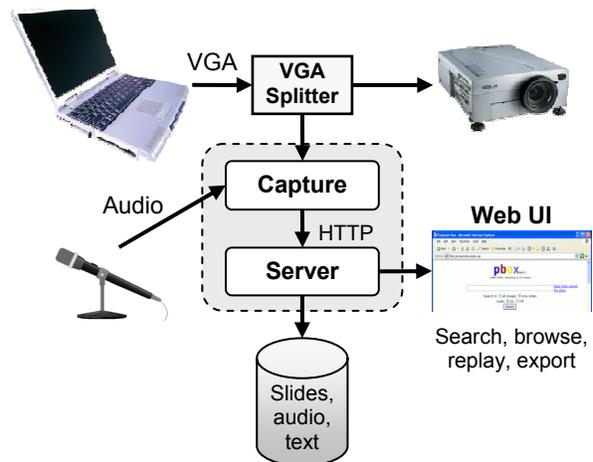


Figure 2. The ProjectorBox architecture.

When the server receives an image, it generates a thumbnail version for the web interface and calls the OCR component to extract its textual content along with the bounding boxes for each word in the image. We currently use Microsoft’s Document Imaging OCR because it can be easily automated by external applications. Finally, the image, text and audio data is time stamped and saved in a relational database. The server also performs slide classification and presentation segmentation (as described above) and provides the web-based user interface for easy retrieval, skimming, and playback.

We use HTTP to transmit images and associated audio clips from the capture component to the server to enable a flexible architecture in which the capture and server components can run on the same or separate PCs. For example, a single server can

integrate content sent from lightweight capture components distributed in multiple classrooms or conference rooms.

One of our goals for ProjectorBox was to minimize storage requirements. On data captured during one year, the average size of one hour of recording is 30MB (250 KB per minute for the MP3, and 400 KB per slide image, with 40 slides per hour). This is ten times lower than state of the art MPEG4 video encoders for similar high-resolution encodings (e.g. 1024x768 pixels).

3.3 User Interfaces

The web interface supports two methods for quickly retrieving content. The main page (Figure 3) shows a list of dates and times in a calendar-like list, indicating when content has been captured. If the date and time of a specific presentation is known, this page provides a single-click solution to presentation retrieval.

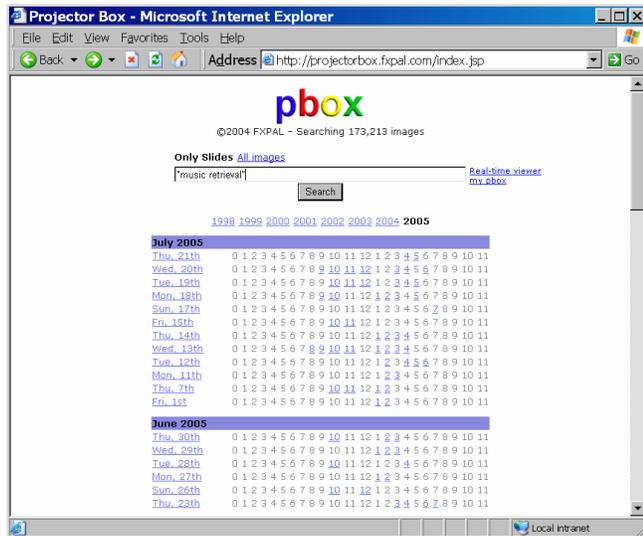


Figure 3. The main page.

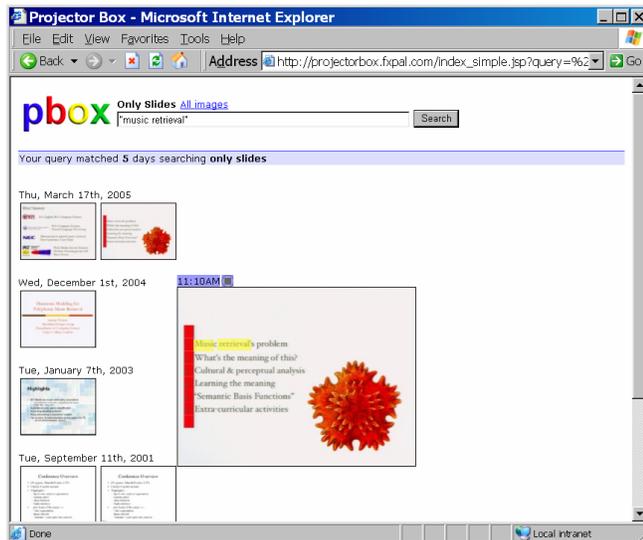


Figure 4. The search results page.

The main page also provides a text field for full-text search of all captured presentations, allowing users to retrieve slides by content. The search results page (Figure 4) shows matching slides organized by date. Moving the mouse over a slide shows an enlarged version with the matching query terms automatically highlighted in yellow, and plays back the associated audio clip.

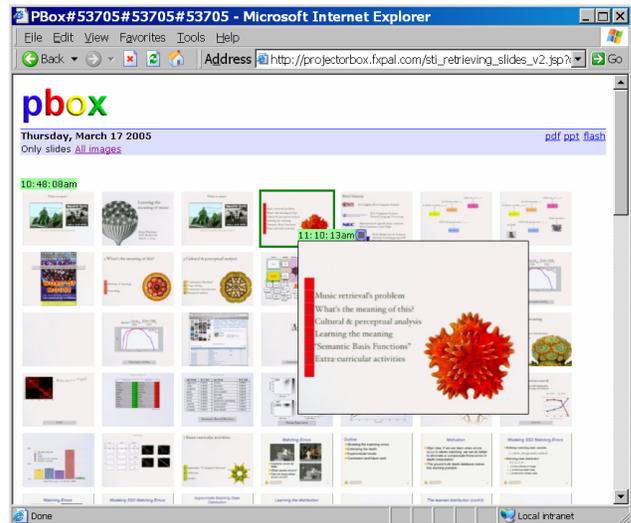


Figure 5. The day viewer page.

When the user clicks on a date or time in the browse interface or clicks on a slide in the search results, the day viewer page (Figure 5) shows thumbnail images of slides captured that day. Here, too, moving the mouse over a slide shows an enlarged version and also plays the associated audio clip. This page makes it easy for users to skim around presentations to locate slides or segments of interest. Double-clicking on an image brings up a slide player for playing back content sequentially or skipping backward and forward through slides.

3.4 Evaluation

We initially focused on capturing static presentation content (namely slide images) and room audio. We reasoned that dynamic content (such as slide animations, videos, and software demonstrations) could be captured later if users wanted it. We also decided to forgo capturing room video for the sake of simplicity and to see whether users would demand it. Would users find the system useful? Would they miss content we hadn't captured?

To find out, we deployed ProjectorBox in our corporate conference room as soon as we had a working prototype. We improved our user interfaces and slide classification and segmentation algorithms based on data collected over the course of several months of use. We then deployed ProjectorBox in two more meeting rooms at Fuji Xerox, and in two university classrooms, one at San Francisco State University (SFSU) and one at the Naval Postgraduate School (NPS) in Monterey. We briefly discuss evaluation results from ProjectorBox use in our corporate conference room over a 35 week period. Results are based on user surveys, web logs, and informal discussions with 22 participants (excluding the authors).

3.4.1 Results

Most users (95%) reported using ProjectorBox to review presentations they were unable to attend. The next most important uses included reviewing presentations users had already seen (85%) and finding out "what's going on in the lab" (75%). Users were far more likely to skim (or playback) portions of presentations than to play them back in their entirety.

Most users (85%) felt more comfortable missing a presentation in the event they were unable to attend. Most users (65%) also reported taking less notes than usual, and felt they were able to

pay more attention in meetings (55%). According to one of our user's, "I don't take notes. I stopped doing that because I never look at them again anyway. ProjectorBox is good, because it not only gives you the image but also what they said. You can get right back to the moment of the presentation and remember what you thought about then."

We also asked our users whether they missed other types of information we were not capturing. Everyone felt it would be useful to capture software demos and video clips, and many felt room video would be useful at least sometimes (65%) or all of the time (20%). In addition to slide images, text, and audio, 70% of our respondents felt it would be useful to be able to search the audio. Finally, about a quarter of users reported having concerns about privacy or security, which we discuss further in Section 5.

3.4.2 Summary

Our results supported our initial vision and design decisions. ProjectorBox was both easy to use and useful. Automatic slide and audio capture helped corporate users review missed presentations and stay informed about "what's going on" within the organization. Our users expressed a desire for more support to find personally relevant presentations they missed and to stay aware of what was going on in the organization. We have since begun investigating new services that leverage ProjectorBox's archive to bring relevant information to users' attention. These include awareness and proactive retrieval interfaces, as well as content-based browsing for information discovery. In the following section we discuss FXPAL's proactive retrieval system.

4. PAL BAR

PAL Bar is a proactive contextual retrieval tool, specifically designed to support seamless access to relevant corporate content. The system was originally designed to support contextual access to corporate and personal contacts, as described in [22]. Since then, PAL Bar has evolved into a much more general information management and recommendation framework [5]. In the following sections we briefly review the system's recommendation interfaces, its architecture, and the recommendation algorithm. In addition, we show how the combination of seamless capture technology, such as ProjectorBox, and seamless discovery technology, such as PAL Bar, has enabled a fully integrated multi-media corporate memory system that supports capture and discovery of content without any added burden on anyone.

PAL Bar's main purpose is to proactively recommend content related to the user's "context", such as a currently displayed web page or email message. The concept of contextual proactive information retrieval is not new (e.g. see Watson [6] and Remembrance Agent [20]), but unlike other systems in this field, PAL Bar was specifically designed for corporate information access: as described in the algorithm outline below, PAL Bar's recommendation algorithm utilizes information about the organization. In addition, PAL Bar supports multiple alternative recommendation interfaces, allowing users to choose between more or less subtle ways of being notified about corporate resources.

4.1 Recommendation Interfaces

PAL Bar is a toolbar for web browsers and is currently available for MS Internet Explorer and Mozilla FireFox. Its initial design was quite simple, consisting of only four buttons and a search text field (Figure 6). The three right-most buttons acted as the

system's first recommendation interface: these *recommendation buttons* are typically grayed out, but change color (from gray to red) when the system recommends resources related to the currently displayed web page. When the user clicks on a red button, the system shows a drop-down menu containing lists of recommended resources. The left-most button produces a drop down menu that allows users to change their preferences about how recommendations are presented, and to browse the full contents of the document database.

Clearly, this interface was subtle and unobtrusive. However, its recommendations were almost universally ignored. Based on a small-scale user study conducted at FXPAL [5], we designed and implemented three additional user interfaces: *translucent recommendation windows*, a *SideBar* interface and an email-based *recommendation digest* that the system periodically sends to PAL Bar users.

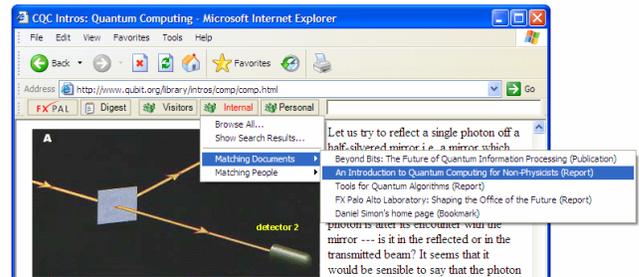


Figure 6: PAL Bar's Initial Recommendation Interface

4.1.1 Translucent Recommendation Windows

Our initial recommendation interface was so subtle that our users tended to forget that it was there. We decided to occasionally display recommendation notification windows in cases in which the system judged a recommendation to be of exceptionally high quality.

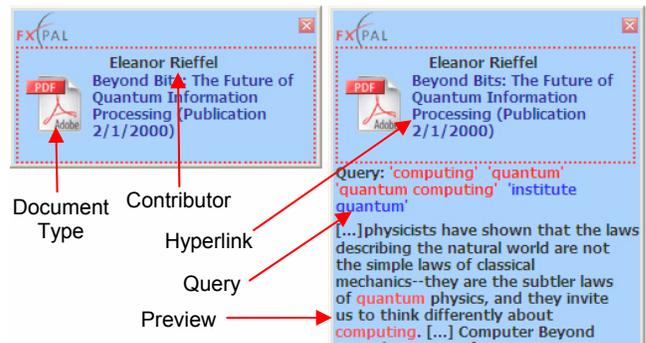


Figure 7: Translucent Recommendation Windows

When a recommendation is judged to be particularly relevant, the server sends a special command indicating that a recommendation window should be displayed. Recommendation windows are small, partially transparent windows that fade in on the corner of the user's screen, similar in style to the translucent email notification windows in Outlook 2003. The windows display a link to the recommended document, the document type, and the generated query terms (Section 4.3), which are colored to indicate which terms did and did not appear in the recommended document. Users can also resize the windows to taste to show a document preview consisting of snippets extracted from the document, with the query terms highlighted. For example, the web page displayed in Figure 6 generated the recommendation window shown in Figure 7. Recommendation windows

automatically fade away after a short delay unless the mouse cursor is moved over them.

There is, of course, a fine line between making our recommendations more noticeable and making them overly invasive or annoying. We addressed this by making our window design small and understated and making the relevance threshold for triggering a window quite high. A follow-up user study at FXPAL [5] revealed that the recommendation windows led to a significant increase in accessed content, but several users stated that PAL Bar had become much more distracting. In addition, the recommendation windows did not address the fact that users may be engaged in other tasks and may therefore not want to interact with the window or recommended content right away. The following two interfaces address this issue.

4.1.2 Sidebar Recommendations

The Sidebar interface is similar, in principle, to the recommendation windows, i.e. particularly relevant recommendations are immediately brought to the user's attention. However, instead of rendering recommendations in a separate window that appears and disappears proactively, recommendations are added to a *Sidebar* interface that remains constantly visible on the right side of the screen. To implement the interface, we used Google Inc.'s popular and freely available *Google Desktop* software (<http://desktop.google.com>), which includes a Sidebar component that can be extended with custom implementations of "panels" that display information within the Sidebar.



Figure 8: Sidebar Recommendation Interface

Figure 8 shows a portion of the Sidebar UI: new recommendations are added to the top of the PAL Bar panel and remain visible until new recommendations cause them to move out of the visible area. When the user clicks on a recommendation, the Sidebar opens a window similar to the one shown in . In summary, this interface does not force users to interact with recommendations right away: recommendations do not fade away after a few seconds, i.e. users can access them whenever they have time to do so. Arguably, the interface is also less obtrusive than the recommendation windows, because the Sidebar only updates its content, but its size and appearance remains unchanged. However, users must give up a small portion of screen real-estate, and in our experience, not all users are willing to do this.

4.1.3 Recommendation Digest

The recommendation digest is different from all three interfaces described above: it does not proactively bring recommendations to the user's attention. Instead, the digest is a document that lists particularly relevant recommendations the user might have

previously missed. By clicking the digest button (see Figure 6), users can request digests that contain a subset of the recommendations presented during the preceding hour, day, or week. Alternatively, users can request that digests be sent periodically via email. Most PAL Bar users currently receive an automatically generated digest document once a week.

To determine which recommendations to include in the digest, we use a recommendation aggregation approach designed to restrict digests to only the most relevant information. This approach is simple and general enough to be applied to a broad range of proactive recommendation scenarios. Whenever a recommendation R is recommended to user U , the pair $R-U$ is logged in our database along with additional information such as the date and time of the recommendation, the context that triggered the recommendation, and a measure of the system's confidence in the recommendation. This produces a recommendation log that can be efficiently turned into a recommendation digest by aggregating individual recommendations. The intuition underlying this process is that items that were recommended multiple times to the same user over a specified period of time (potentially based on multiple different contexts) are more likely to be related to the user's information needs or interests than items that were recommended infrequently or only based on a small number of visited pages.

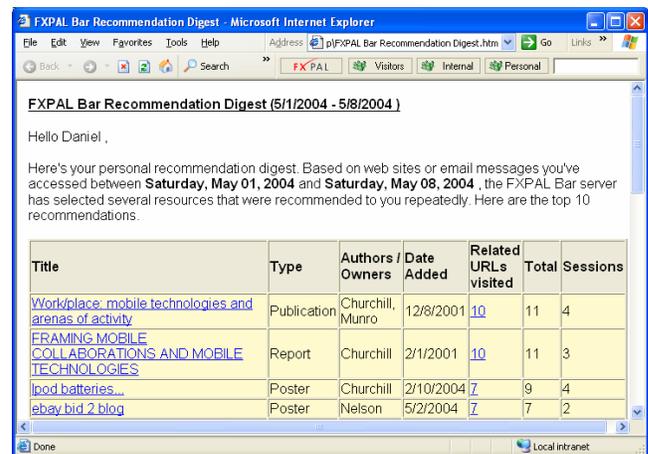


Figure 9: Recommendation Digest Example

The main goal of the aggregation process is to identify recommendations that were repeatedly recommended to the same user during a specified period of time, preferably based on a large number of distinct contexts. The server generates the following statistics as part of the aggregation process:

- *Number of distinct contexts.* In the example shown in Figure 9 this is the number of related URLs visited, and all listed documents are sorted by this number. The number of distinct contexts is likely to be a strong relevance indicator, as every recommendation of the same document based on a different context can be interpreted as an additional piece of evidence supporting the document's relevance.
- *Total number of recommendations.* This is the total number of times a document was recommended, including recommendations based on repeat visits to the same page.
- *Number of distinct sessions.* This is the number of distinct usage sessions during which a resource was recommended, which is an additional indicator of recurring long-term interests.

The example digest shown in Figure 9 focuses on the user's most significant browsing patterns over a week and brings relevant corporate resources to the user's attention. In the example shown, the digest is rendered as a web page that the user has explicitly requested. The same digest can optionally be emailed to users at user-specified intervals.

4.2 System Architecture

PAL Bar uses a client-server framework in which all Bar clients are powered by a single server (Figure 10). This server exposes a set of web services that support various content management and recommendation functions, allowing it to be accessed from any software component that can interact with web services. This allows external applications or other embedded clients, such as toolbars for email applications or word processors, to take advantage of the system's recommendation capabilities.

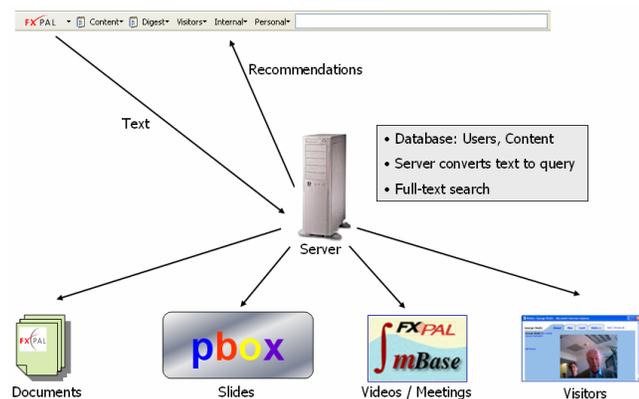


Figure 10: PAL Bar Architecture

The server runs a relational database that supports full-text search, which it uses to maintain information about users and available content. As part of FXPAL's corporate memory system, the server has access to multiple content repositories, such as:

- Documents from PALWeb, FXPAL's Intranet, such as reports, memos, publications or patent applications (see Section 2.1).
- Presentations captured by ProjectorBox (see Section 3).
- Videos captured by UbiSight (see Section 2.4), stored in FXPAL's multimedia database (mBase).
- Web pages posted to our Plasma Poster (see Section 2.3).
- Visitor information from FXPAL's visitor guestbook, containing visitors' contact information, recorded videos, and uploaded documents, as described in [22].
- Any other information shared by individual users: document collections, web bookmarks or contact lists.

To keep the system's content up to date, we automated the content import process as much as possible. For example, the system periodically imports presentations captured by ProjectorBox and UbiSight, as well as documents published on PALWeb. In addition, when users bookmark new web pages, the system can automatically (or interactively) add the page to its document repository.

The information flow through the system is straight-forward: when a user navigates to a web page, PAL Bar extracts the full text of the page (currently ignoring any layout or format information), and sends it to the server. The server then identifies

matching resources and sends a set of content recommendations back to the client. The client then renders the recommendations using one or several of the interfaces described in Section 4.1.

4.3 Recommendation Algorithm

PAL Bar currently supports two separate recommendation approaches: (1) *contact* recommendations based on explicit matches of known contacts in the currently displayed page, and (2) *content* recommendations based on textual similarity between the currently displayed page and other documents in the system's database.

To determine contact matches, the server's recommender component analyzes the transmitted text and uses a matching algorithm to detect occurrences of contact information fragments that match entries in the system's contact database. The algorithm is sophisticated enough to deal with a wide variety of potential formatting differences of names and contact information [22].

Recommendations based on content similarity are determined via a simple two step process. First, the query generation step converts the currently displayed web page to a weighted query. Second, the recommendation step uses the query to determine a set of candidate documents and then determines whether the retrieved candidate documents should be recommended.

The query generation step proceeds as follows. When the server receives the extracted text of the currently displayed web page, it is converted into two separate normalized Vector Space Model *tf-idf* term vectors [23]. One term vector uses individual words (unigrams), while the other term vector uses word pairs (bigrams). The underlying *df* component of these term vectors is based on the server's document collection and an additional set of anonymously logged, previously visited web pages. This ensures that the document frequencies reflect both the document collection and users' actual browsing patterns.

The unigram and bigram vectors are sorted by their respective term weights, and the server uses the resulting term lists to construct the query in three steps. First, up to n unigrams that exceed a fixed threshold t (currently set to 5 and 0.2, respectively) are taken as the initial query terms. Second, up to m bigrams that exceed another threshold u are added to the query (currently set to 5 and 0.1, respectively). Finally, the third query generation step biases the query toward the organization's document collection. The underlying intuition is that our knowledge about the available documents and recurring topics within these documents can be used to add query terms that represent important topics within the document collection, even if they do not stand out statistically on the currently displayed page. For example, several researchers at FXPAL have done work on video segmentation, and as a result, FXPAL's document collection contains many documents that contain the bigram "video segmentation." The goal of the final query generation step is to include bigrams like "video segmentation" if they appear on the currently displayed web page, regardless of their associated *tf-idf* weights, thus making it more likely that highly relevant aspects of a displayed web page contribute to the query. In our implementation, this step requires a pre-computed list of informative query bigrams, which is currently generated once a week according to the following algorithm. All documents in the document collection are converted to *tf-idf* bigram vectors, and restricted to the n top-ranked bigrams. For each bigram in the resulting vocabulary, the server then counts the number of times the bigram occurs in a top- n bigram vector, and sorts bigrams according to this frequency

count. The top m bigrams from this list form the informative query bigram list. The server then uses all unigrams and bigrams identified in the first three steps and constructs a weighted query, where the weights are normalized $tf-idf$ weights of all query terms.

The recommendation step proceeds as follows. First, the server retrieves the full text of the top n query results and constructs corresponding $tf-idf$ unigram vectors (n is currently set to 10). The server then determines the exact similarity of the current web page and each retrieved document using the cosine similarity measure [23], and sorts the documents accordingly. Documents that exceed a similarity threshold t are recommended to the user.

4.4 Seamless Media Capture and Discovery

This section briefly describes how PAL Bar, when integrated with seamless capture technology such as ProjectorBox and UbiSight, leads to a compelling end-to-end corporate memory solution that supports seamless media capture and discovery without any added burden on anyone. The scenario described in this section exemplifies the core idea that is driving our research on corporate memory and automated information dissemination: knowledge capture and reuse as byproducts of established everyday work practices.

As described in Section 2.4, *UbiSight* is a system that automatically records videos of presentations or meetings, based on motion and audio tracking. Since *UbiSight* does not capture any textual content, retrieval of automatically recorded videos is challenging. However, *ProjectorBox* addresses this problem: since *ProjectorBox* stores the exact time of every recorded slide, the resulting index can be used to locate video segments that correspond to displayed slides. In short, *ProjectorBox* serves as a full-text slide and video index.

This means that in addition to textual documents, PAL Bar is now able to recommend automatically recorded meetings and presentations. For example, consider a scenario where a researcher at FXPAL visits a web page about “collaborative filtering”. Based on the text of this page, PAL Bar proactively generates a query that retrieves a set of corporate documents related to this topic. Figure 11 shows how the Sidebar interface renders such a set of recommendations. The first recommendation in this set is a presentation that was automatically recorded by *UbiSight* and *ProjectorBox*. In the scenario shown in Figure 11, the user selected this presentation, causing the Sidebar to display a recommendation window that allows the user to browse and access individual slides.

Since the system also has access to corresponding *UbiSight* video, the user can start FXPAL’s video and slide player to play back the recommended presentation (starting at the currently displayed slide). Figure 12 shows FXPAL’s corporate memory media player: video is played back on the left side of the window, while corresponding slides are shown on the right side. The timeline at the bottom of the window allows users to see slide transitions (vertical bars) and to navigate to any slide. Slides that match the user’s original context, in this case a web page on “collaborative filtering” are considered particularly relevant and are marked with bold squares above the timeline.

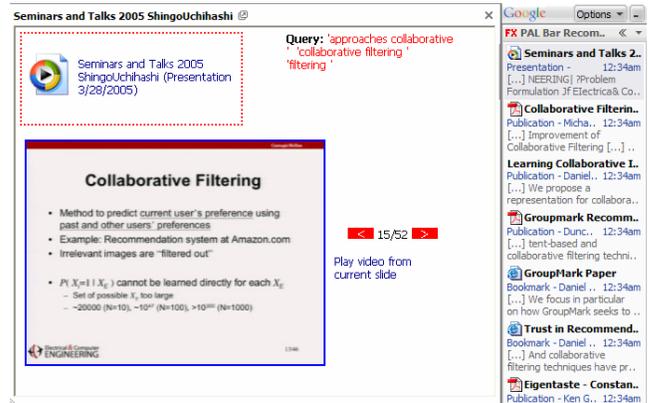


Figure 11: Sidebar and a Selected Recommendation

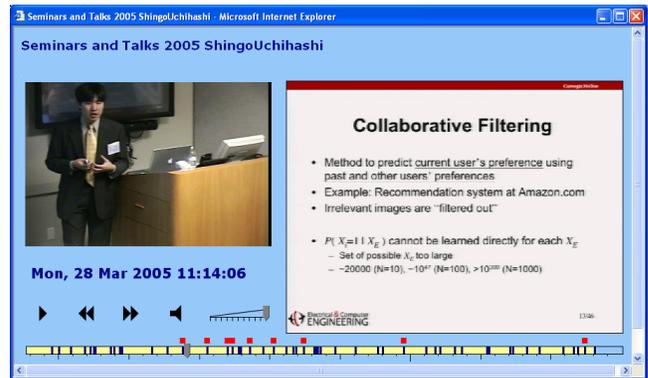


Figure 12: Corporate Memory Media Player

In summary, the integration of *PAL Bar*, *ProjectorBox* and *UbiSight* enabled a powerful corporate memory system for completely seamless information capture and discovery. While a formal evaluation of the utility of the system integration will be the subject of future work, we feel that we are one step closer to realizing Vannevar Bush’s Memex vision [7] for organizations: we have built a fully automated “Corporate Memex”.

5. FUTURE WORK

Our future work will focus on extensions that will make our corporate memory system more complete, secure and universally accessible. We have implementation plans for near-seamless email capture that supports user awareness and consent, and does not infringe on our users’ privacy. Related to this, we will make the system more secure, by adding user-friendly access control solutions. Finally, we plan to make our corporate memory content more accessible via an extensible knowledge portal that provides integrated access to all manually or automatically acquired content. The following sections briefly outline our current research and development in these areas.

5.1 Email Capture

Employees of any organization commonly exchange valuable information via email. For example, employees may discuss ideas, ask and answer questions, comment on corporate activities, or share and discuss documents. The vast majority of these messages is sent to individuals or small groups of recipients. However, a short recipient list does not necessarily mean that the message contains confidential information that cannot be shared with others. For example, email authors may decide to send information only to their teammates, select recipients based on an

assessment of how immediately relevant the content may be to the recipient, the likelihood of getting an answer to a question from a presumed expert, or simply the degree of familiarity with a recipient. Similarly, potential recipients may be excluded from messages simply because the sender does not want to bother people who, presumably, may not be immediately interested. As a result, information and content communicated via email is often not effectively utilized, because employees who were not on the original list of recipients have no way to search for or discover the communicated information.

For privacy reasons, making email globally accessible or searchable is clearly not a viable approach to corporate knowledge sharing. Based on this problem, we investigated and designed techniques that allow organizations to better utilize email content without infringing on established email practices, or sacrificing employees' privacy and control over shared content.

Our approach consists of methods that allow employees to easily express email sharing permissions, which are subsequently used to share content on-demand. In addition, the granted permissions are communicated to all original email recipients and authors to enhance awareness of sharing decisions made by individual employees. The following paragraphs provide a brief overview of the four main components that enable seamless email capture with user awareness and consent, and briefly outline how the captured content can be utilized.

Sharing Designation. Authors must be able to explicitly assign access privileges in a way that does not cause significant usage overhead and does not interfere with established email practices. A convenient and near-seamless way to label messages is to "carbon copy" a virtual email address that represents the group of employees allowed to access the message. For example, CCing an account named cm (short for "corporate memory") may allow employees to assign company-wide access privileges to the message with only two extra key strokes.

Sharing Notification. Since our approach enables near-seamless email capture, it is particularly important to support explicit notification and control mechanisms, so that authors retain control over their content, even if it was shared by others. This can be accomplished in two ways: message modification and message generation. Message modification based on email interception supports our goal of supporting established email practices without burdening the user with any extra overhead. To increase awareness of implicitly designated access rights, our system will intercept and modify all messages that are sent to shared email addresses. For example, the system may modify the subject and/or body of the message to clearly indicate that the message is accessible to a wider group of people than the immediate recipients. In addition, the modification will include a link that takes the user to a content management interface, so that authors can remove previously shared messages from the corporate memory, if necessary. Since the system intercepts and modifies messages before recipients receive them, this approach does not generate any extra message traffic or overhead for the user.

In addition to message interception and modification, our approach includes notification mechanisms designed to communicate sharing decisions that were made without explicit consent of the author. For example, someone may decide to forward someone else's message or an entire email thread to a designated "shared" email address without including the author as a message recipient. In this case, the system can generate a new

message to notify the author that content was shared and optionally provide the means to edit or remove the shared content.

Content Processing. Our system stores and indexes shared messages, along with embedded and linked content. This means that, in addition to the full text of the message subject and body, the system separately indexes attached files (e.g. Word or PDF files), and also downloads, stores and indexes linked web pages and documents. The system also retains information about the origin of indexed documents, i.e. an attached or linked document can always be traced back to a particular email message. This provides valuable context for employees: e.g., a search result would not only include a matching document, but also provide the document's context, such as the email message that originally included the document.

5.2 Privacy and Security

Clearly, any attempt at automating the acquisition and utilization of corporate content has important privacy and security implications. Promoting user awareness, consent, and control while maintaining extreme ease of use is challenging.

Below we discuss how privacy and security concerns have arisen in the context of ProjectorBox, and how we are addressing them. Clearly, some of the described concerns are not specific to ProjectorBox and affect other components of our system. For example, user awareness, consent, and control guided our design of the email capture features described in Section 5.1.

When employees or visitors present in our corporate conference room, they must be aware of, and consent to, being recorded. In the past, we simply notified presenters before recording them. However, since ProjectorBox records presentation without anyone having to remember to start recordings, we have forgotten on occasion to notify presenters ahead of time. Clearly, this social approach, which used to work before ProjectorBox, is no longer adequate. We have since added prominent red warning labels to the podium PC (used for most presentations) and the VGA cable connector (used by visitors to connect their laptops) to inform presenters they will be recorded. Other issues include the ability to easily suspend recordings (for "off-the-record" comments or to prevent sensitive meetings from being recorded) and access control over captured content.

Ideally, ProjectorBox would overlay a visible recording indicator in the projected video stream so presenters and audience members would know when a recording was in progress. ProjectorBox could also detect and analyze the presenter's mouse movements in the video stream and recognize gestures around the overlaid recording indicator to provide a simple UI for presenters to stop and start recordings. This approach is technically feasible and has the virtue of not requiring presenters to interact with additional interfaces separate from the interfaces they naturally use for presenting.

Another approach that works well when presenters handle the VGA cable involves embedding a recording indicator and switch directly into the VGA cable connector so presenters notice and interact with it at the time they connect their laptop. Such a "smart cable" would indicate (e.g., using LEDs) whether recording was activate or inactive and provide a physical switch to allow presenters to easily opt-in or opt-out.

Finally, to address access control issues, we have begun experimenting with smart cards for user identification and alternative form factors for the capture device itself. Smart cards

can be used to identify presenters to ProjectorBox, which can initially keep captured content private, and then email a link allowing the presenter to review content before releasing it to others. Alternatively, if users can capture their presentations using an ultra-portable personal (or workgroup) device, then privacy and access control issues become less problematic, since the user (or team) has physical control over the captured content and can set access controls when they upload it to a server.

5.3 Enterprise Knowledge Portal

Since access to some corporate resources is currently limited to PAL Bar's search and recommendation UI (see Section 4), we are currently implementing an enterprise knowledge portal that provides integrated web access to all manually and automatically acquired content. The implementation uses the popular *Portlet* [14] standard, allowing individual developers in our lab who are in charge of individual corporate memory components to contribute to our portal. In addition to integrated full-text search, we plan to include Portlet versions of some of the interfaces discussed in this paper. For example, the recommendation digest discussed in Section 4.1.3 is currently being converted into a Portlet, which means that our employees will be able to configure the portal front page to automatically display corporate resources related to recently visited web pages.

6. CONCLUSION

This paper described how two key technologies, ProjectorBox—a smart appliance for automatic presentation capture—and PAL Bar—a system for proactively retrieving contextually relevant corporate memories—have enabled us to integrate content from a variety of sources to create a cohesive multimedia corporate memory for our organization. The system integration and usage scenarios described in this paper exemplify the core ideas that are driving our research on corporate memory and automated information dissemination: knowledge capture and reuse as byproducts of established everyday work practices. While a formal evaluation of the utility of the resulting solution will be the subject of future work, we feel that we are one step closer to realizing Vannevar Bush's Memex vision [7] for organizations: we have built a fully automated "Corporate Memex".

7. REFERENCES

- [1] Abowd, G.D., Classroom 2000: An experiment with the instrumentation of a living educational environment. *IBM Systems Journal* 38, (1999), 508-530.
- [2] Anystream. *Apreso Classroom*. <http://www.apreso.com/>.
- [3] AutoAuditorium. <http://www.autoauditorium.com/>.
- [4] Autonomy. <http://www.autonomy.com/>.
- [5] Billsus, D., Hilbert D, and Maynes-Aminzade, D.: Improving Proactive Information Systems. In Proc. IUI 2005, (2005).
- [6] Budzik, J., Hammond K., and Birnbaum, L. Information Access in Context. *Knowledge-Based Systems* 14 (1-2), (2005), 37-53.
- [7] Bush, V. As We May Think. *The Atlantic Monthly*, July, (1945), 101-108.
- [8] Chiu, P., Kapuskar, A., Reitmeier, S., and Wilcox, L., Room with a Rear View: Meeting Capture in a Multimedia Conference Room. *IEEE MultiMedia Magazine*, 7, 4, (2000), 48-54.
- [9] Churchill, E.F., Nelson, L., Denoue, L., Helfman, J.I. and Murphy, P. Sharing Multimedia Content with Interactive Displays: A Case Study. In Proc. DIS 2004, ACM Press (2004).
- [10] Cruz, G. and Hill, R. Capturing and playing multimedia events with STREAMS. In Proc. MULTIMEDIA '94, ACM Press (1994), 193-200.
- [11] DataPath. <http://www.datapath.co.uk/visRGBPRO.htm>.
- [12] Erol, B., Hull, J. J., and Lee, D. Linking multimedia presentations with their symbolic source documents: algorithm and applications. In Proc. MULTIMEDIA '03, ACM Press (2003), 498-507.
- [13] Hilbert, D.M., Cooper, M., Denoue, L., Adcock, J. and Billsus, D. Seamless presentation capture, indexing, and management. In Proc. SPIE Optics East, (2005).
- [14] Java Portlet Specification (JSR 168). JCP specification, October 2003. <http://jcp.org/aboutJava/communityprocess/final/jsr168/index.html>.
- [15] Liu, Q., Rui, Y., Gupta, A., and Cadiz, J. J. Automating camera management for lecture room environments. In Proc. CHI '01, ACM Press (2001), 442-449.
- [16] Liu, Q., Shi, X., Kimber, D., Zhao, F., Raab, F. An Online Video Composition System. *IEEE International Conference on Multimedia & Expo*, (2005).
- [17] Minneman, S., Harrison, S., Janssen, B., Kurtenbach, G., Moran, T., Smith, I., and van Melle, B. A confederation of tools for capturing and accessing collaborative activity. In Proc. MULTIMEDIA '95, ACM Press (1995), 523-534.
- [18] Mukhopadhyay S. and Smith, B., Passive capture and structuring of lectures. In Proc. MULTIMEDIA '99, (1999), 477-487.
- [19] Quindi. <http://www.quindi.com/>.
- [20] Rhodes, B. Just-In-Time Information Retrieval. Ph.D. Dissertation, MIT Media Lab, 2000.
- [21] Sonic Foundry. *MediaSite*. <http://www.mediasite.com>.
- [22] Trevor, J., Hilbert, D. M., Billsus, D., Vaughan, J., and Tran, Q. T. Contextual contact retrieval. In Proc. IUI '04. ACM Press, (2004), 337-339.
- [23] Van Rijsbergen, C. J.: *Information Retrieval*. London: Butterworths, (1979)
- [24] Verity. <http://www.verity.com/>.
- [25] Walsh, J.P. and Ungson, G.R. Organizational Memory. *Academy of Management Review*, 16, 1, (1991), 57 - 81.
- [26] Weiser, M. The computer of the 21st century. *Scientific American*, 265, 3, (1991), 66-75.
- [27] Ziewer, P., Navigational Indices and Full-Text Search by Automated Analyses of Screen Recorded Data. In Proc. of E-Learn 2004, (2004), 3055-3062.