Design and Implementation of a Historical Database

for the Blacksburg Electronic Village
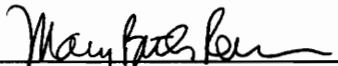
by

Kenneth William Schmidt, Jr.

Research project report submitted to the Faculty of the Virginia Polytechnic Institute and

State University in partial fulfillment of the requirements for the degree of
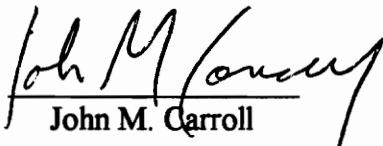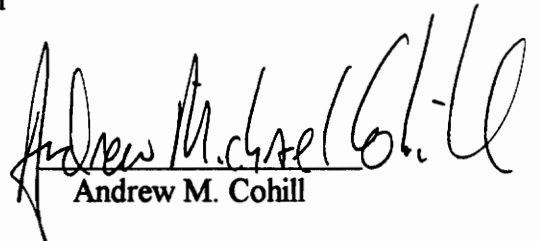
MASTER OF SCIENCE

IN

COMPUTER SCIENCE

APPROVED:

Mary Beth Rosson, Chair

John M. Carroll

Andrew M. Cohill

February, 1996

Blacksburg, Virginia

Key words: History, Documentation, Community, Networks

C.2

Design and Implementation of a Historical Database

for the Blacksburg Electronic Village

by

Kenneth William Schmidt, Jr.

Mary Beth Rosson, Chair

Computer Science

(ABSTRACT)

This research project report discusses the development of a dynamic, self-documenting, historical database (HistoryBase) of the Blacksburg Electronic Village (BEV). The research project consists of two parts, the development of a system to be accessed and extended via the World-Wide Web, and a brief analysis of the issues faced during the design and implementation of the system.

Table of Contents

List of Figures

List of Tables

Problem Description

Over the past few years, hundreds of cities and towns around the world have either started or plan to start building community networks [CityLink96, Morino96]. This is due in part to the growing popularity of the Internet and the World-Wide Web and to the press coverage given to a few communities who are pioneering these type of networks [BEV96, Boulder96].

Documenting the design of such projects could be of great use to others trying to build their own community networks. Many of these newer networkers will look to the pioneers in this area for guidance - to learn from past successes and mistakes and to discover issues that may not have been considered. In the early stages of their projects, these networkers may want to find strategies for soliciting funding from government agencies and local utilities and also successful methods for promoting the project. In later stages they may look for ways of educating the public about community networks and encouraging local businesses to participate.

Not only will developers of other community networks find this useful, but also documenting what happens as a community network evolves could be of interest to its users, as it may contain historical information about community members, groups and activities. For example, a parent may want to learn more about the events that led up to

the first electronic PTA meeting, or a retiree may want to see who was instrumental in creating the senior's presence on the network.

What is the best way to create this documentation? By definition, a community network is a sprawling, distributed activity, with no one central authority figure. Furthermore, even if you tried to get project developers to document it, it is a well-known that designers are not very good at producing documentation; they think it is tedious and are not motivated to produce material that would only benefit someone else [Weiss83].

The current project seeks to find a viable way to document such a project in a collaborative and distributed way -- addressing the tediousness and difficulties of documenting a large-scale system by involving the entire community. It leverages the fact that the network is part of the World Wide Web, automating the collection of information, and integrating collection of additional information with use of the community network itself.

This document is organized in the following manner: this section (Section 1) provides a brief overview of the problem, Section 2 describes the design and implementation of the BEV HistoryBase, Section 3 analyzes some of the important issues faced during system development, and Section 4 suggests directions for future work.

Project Description

What is History?

According to Webster's Collegiate Dictionary, history is "a chronological record of significant events... often including an explanation of their causes [Webster94, p. 550]." Landes and Tilly define history as "the branch of inquiry that seeks to arrive at an accurate account and valid understanding of the past [Landes71, p. 5]." Vincent discusses the history of the notion of history itself. Primitive history was told as stories, with "no questions... asked about the theories upon which it was based, nor concerning the motives which actuated the narrator [Vincent74, p. 1]." Vincent notes that the work of Herodotus is the first known example of "extended historical composition," where Herodotus states that he wishes not only to record the events, but also to give reasons for them as well.

A modern history is a record of a set of events along with an interpretation of these events. This project helps to create a record of the events that occur in the Blacksburg Electronic Village, so that others can use this record as a base for analysis of the BEV itself. The project also supports interpretation of documents or events through a document annotation system

Historical Databases

Much of the current research in historical databases has been in the area of design history and group or organizational memory [Atwood95, Cutkosky94, Terveen95]. Each of the systems reviewed here attempt to decrease the effort involved in information seeking in software development organizations.

Answer Garden

Answer Garden [Ackerman94] is a tool for collecting and managing organizational memory. It supports this task by making knowledge retrievable and making individuals with knowledge accessible. The system was designed to capture recurring questions and their answers in a central database. To find information in Answer Garden, the system guides the user through a branching network of questions. The user continues until the question that needs answering is found, where they receive an answer. The response might include text or graphical images, or it may be an "active node" which performs a certain action at run time (e.g., querying a database).

If the answer is not found or is incomplete, the user may use the system to ask a question and this is routed to the appropriate human expert. The expert then answers the question via electronic mail and can put the answer back into Answer Garden. By this mechanism Answer Garden's database grows with system usage.

Designer Assistant

Designer Assistant (DA) was developed to manage design knowledge for a large software development organization [Terveen95], and is similar to Answer Garden in its approach. This system attempts to capture the knowledge that is typically not written down -- knowledge previously maintained only in the minds of the experts and passed from person to person by word of mouth (Terveen, et. all refer to this type of management as "folklore").

DA supports two kinds of "knowledge evolution" -- knowledge update, which is refining and elaborating the existing system knowledge base, and knowledge addition, which is adding new knowledge as it is generated during systems development activities. DA also supports knowledge usage, where developers can access information in the knowledge base and contact the person responsible for encoding that knowledge (the "owner" of that knowledge). Note that in DA the owner of a particular piece of knowledge can change as people leave the organization or their responsibilities shift.

Like Answer Garden, DA asks the user questions to step through a decision tree until the answer is found, or the expert associated with that knowledge has to be contacted because the answer is not there. DA also allows the user to enter more detailed comments and suggestions after the answer is found.

Design Intent

Design Intent [Atwood95] is a system that facilitates communication amongst stakeholders in team software development projects. In this system communication artifacts - project documents, release notes, meeting minutes, etc. - are captured and made available for stakeholders to comment, critique, and/or clarify. One implementation of Design Intent uses the World-Wide Web.

Stakeholders are allowed to identify which artifacts of their communication they want to include, and in what format (textual, audio, video). These artifacts serve as seed documents into the Design Intent system. Collaboration is facilitated through two methods: annotations and bulletin boards. Annotations are a means of commenting, critiquing, and/or clarifying Design Intent documents. Bulletin boards are similar to newsgroups in that discussion threads can develop on a certain topic. While these bulletin board topics need not be related directly to a particular document, they may eventually lead to the creation of a new document.

The system can be accessed through browsing or searching, and new additions are announced to members through either electronic mail notifications, newsletters, or personalized news pages. Also when a comment is made on a Design Intent document the author of that document is notified so that s/he can respond to that comment.

Approach

For this project a historical database system (referred to in this document as the "BEV HistoryBase" or "HistoryBase") has been developed for the Blacksburg Electronic Village. This system records project-related documents such as electronic mail messages, meeting minutes, press coverage, and newsgroup postings by members of the BEV. The HistoryBase system, while developed specifically for the BEV, can be easily extended to other projects or systems where recording historical data is of benefit.

The BEV HistoryBase

The BEV HistoryBase is a hypermedia history system that is used to integrate the collection, organization, interpretation, and dissemination of documents and other materials pertaining to the Blacksburg Electronic Village. The system is accessed using the World-Wide Web, and is linked to the other Web pages that make up the BEV Web site.

Design

User Classification

Before creating any software system it is necessary to identify who will use the system and how they will use it. For the HistoryBase, three different user groups have been identified:

BEV project personnel, the Blacksburg community, and the developers of other community networks [Carroll95]. A brief profile of each of these groups follows.

BEV project personnel

The BEV project personnel are the core group of administrators and technicians who have direct and immediate influence over the BEV, present and future. This group is expected to contribute most of the documents in the HistoryBase. For the system to be accepted by BEV project personnel the benefits need to be readily apparent. If the system is seen as bothersome, not easy to use, or just not relevant, project personnel will not use it.

BEV personnel might use the HistoryBase to record internal documents such as vision statements and software documentation; external documents including BEV brochures and user's guides; events such as demonstrations and community-outreach programs; and views of the BEV including clippings of press coverage and interviews with project personnel and BEV users.

Blacksburg community members

Blacksburg community members make up a much larger group of people than the BEV project personnel. This group has a less direct influence over the BEV than the project personnel -- they are not immediately responsible for project decision-making, but their ideas and opinions can influence project personnel.

Blacksburg community members might contribute to the system by commenting on existing HistoryBase documents or submitting short stories or essays about their experience with the BEV.

Developers of other community networks

The developers of other community networks will not contribute documents to the HistoryBase; however, they may investigate the information contained in the system for answers to the questions that arise in building their own networks. For example they may look in the HistoryBase to get an idea of the costs of routing a high-speed data line to an apartment complex.

Collaboration in the HistoryBase

The HistoryBase is a collaborative authoring system, but not at the document level. Collaboration in the HistoryBase means that the set of documents is created through the joint efforts of many individuals in the BEV project group and community. The HistoryBase does not directly support multiple authoring of a single document. The focus of the system is to support the *collection* and *organization* of existing documents, not to support document authoring. The authoring support is minimal and is focused on the annotation of documents that have been submitted by other users.

In the HistoryBase, each document can have only one contributor. Documents can be created in collaboration outside the HistoryBase and then submitted by one of the

collaborators (or a third party) after the document is finished. A document in the HistoryBase cannot be edited or revised. To update a document, an author would need to submit a new version of the document as a separate object, or add an annotation to the existing document. This decision was made to ensure consistency between annotations to a document and the document's content. If a comment is made on a document, and that document is later changed, the comment may no longer make sense.

Data model

The basic unit of storage in the HistoryBase is a main document. A main document can be one of many different digital media types: text, audio, image, video, or other (see Figure 1). A textual document can be of many formats, from a project design document to a description of a BEV event. An audio or video document can be anything from welcome messages by BEV personnel to clips of press coverage of the BEV. Images can be real pictures or computer-generated graphics.
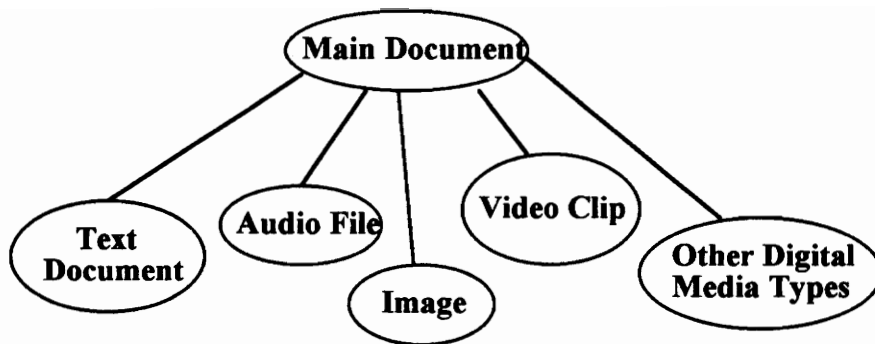
Figure 1: Main Document data types

Meta-information about each main document is also stored in the HistoryBase (see Figure 2). The title is stored separately so that it can be easily accessed and used as an identifier for that document. For documents that describe an event occurring over a period of time (e.g.: installation of communications lines in an apartment complex), the begin and end dates mark the time period. If the document is not time-oriented the begin date will be used to represent when the document was written or released (e.g.: BEV Purpose Statement, February, 1994). When a document is contributed to the HistoryBase, the submit date is set by the system. The classifications are tags that the contributor can give to a document when it is submitted.

```
┌─────────────────────────┐
│ Main Document           │
├─────────────────────────┤
│ Title                   │
│ Body                    │
│ PID                     │
│ Longname                │
│ Event Begin Date        │
│ Event End Date          │
│ Submit Date             │
│ Classification(s)       │
│ Annotation(s)           │
└─────────────────────────┘
```

Figure 2: Main Document meta-data

The PID and the Longname refer to the personal identifier and the real name of the person

who contributed the document at the time it was submitted. Because a single person can

contribute multiple documents, traditional database normalization rules call for the

contributor information to be stored in a separate database table with a link from the main

document record to the Contributor record (see Figure 3). This way if a contributor's

information changes (name change or PID change) only one record will need to be

changed, not every main document record in the database that this person contributed.

The HistoryBase was not designed this way because the intent is to keep an accurate

record of the historical documents and events of the BEV. The history of the people who

contribute the documents is just as much a part of history as documents themselves. In

the HistoryBase if Jane Doe contributes a document in February, 1995, and changes her

name to Jane Smith before contributing a document in May, 1996, the HistoryBase will

reflect Jane's name *at the time she contributed each document* instead of only recording

Jane's current name.
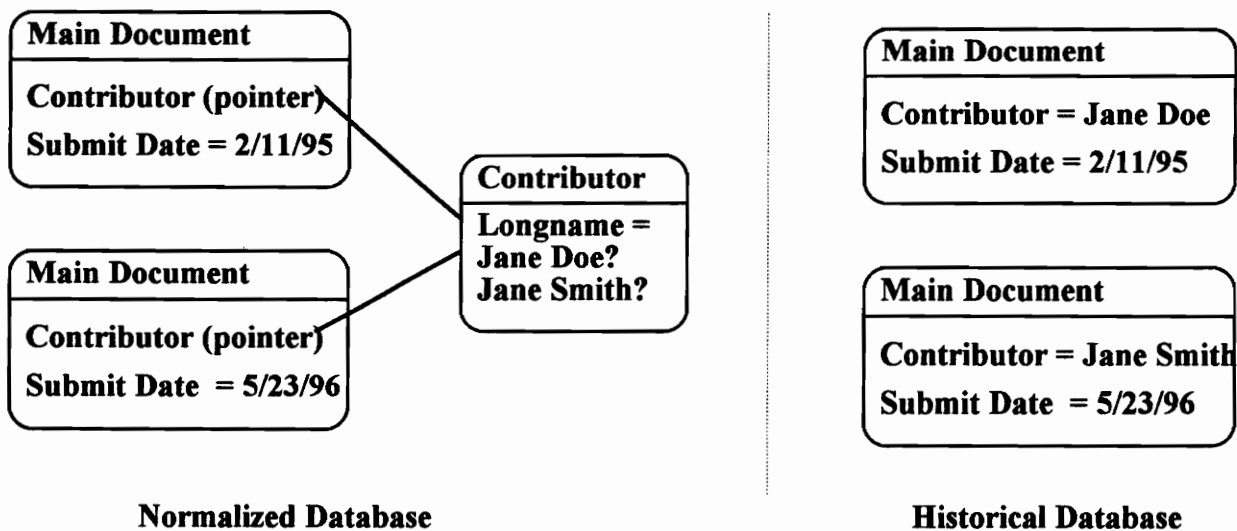
Figure 3: Normalized v/s Historical Database

Finally, every main document in the HistoryBase will be open to comment by any BEV

user. The relationship between a main document and a comment is one-to-many; a single

main document can have many different comments. These comments will be displayed

below the main document in order from least-recent to most recent. This ordering is used

so that the flow of discussion can be seen from the original document through all the

comments on that document.

Like main documents, every comment in the system has some meta-information stored

with it (see left portion of Figure 4). Each comment has a title or subject which

summarizes the content of the comment. Information about who made the comment and the date the comment was submitted is stored also. Finally there is a link from the comment to the main document to which this comment pertains (Comment On). Ordering of comments can be inferred from the submit date. For example in the right portion of Figure 4, both comment X and comment Y are related to the main document, but since comment Y was submitted after comment X, it will be displayed after comment X on the Web page. This comment system allows many different perspectives on the same document to be shared amongst all users of the HistoryBase. In this sense the original document takes on a collaborative nature, with comments or reactions provided by any number of other users (including the original author).
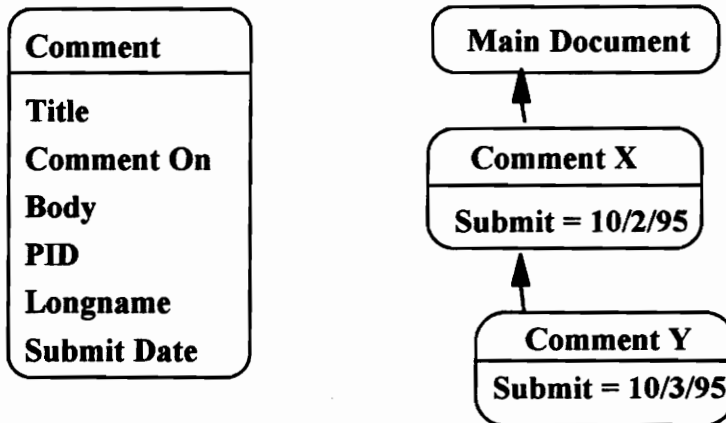
Figure 4: Comment subsystem

Structure of the data space

All main documents, their comments, and the meta-data for each are stored in a database management system (DBMS). This allows flexibility in the type of structure that can be placed around the documents. Any attribute of this meta-data can be accessed and used as an index. For example all documents in the HistoryBase can be organized by a author, thus giving the ability to study an individual contributor's perspectives on the BEV.

Views of the data space

The main interface to the HistoryBase uses a timeline metaphor with landmark events (see Figure 5). This metaphor is used because of the common use of timelines to organize historical events. The landmark events help to give anyone using the HistoryBase a bird's-eye view of the events that are a part of the history of the BEV.

15

BEV HistoryBase: Main Timeline

Click in a box to see a more detailed history for that quarter

First press conference: Partnership formed
between Blacksburg, Tech, & Bell

First apartment
with ethernet

Visit by Lieutenant
Governor

Official Opening
(October 25)

NBC Nightly News
story (Feb 10)

Design History
project begins

Business Grants
for Web Sites
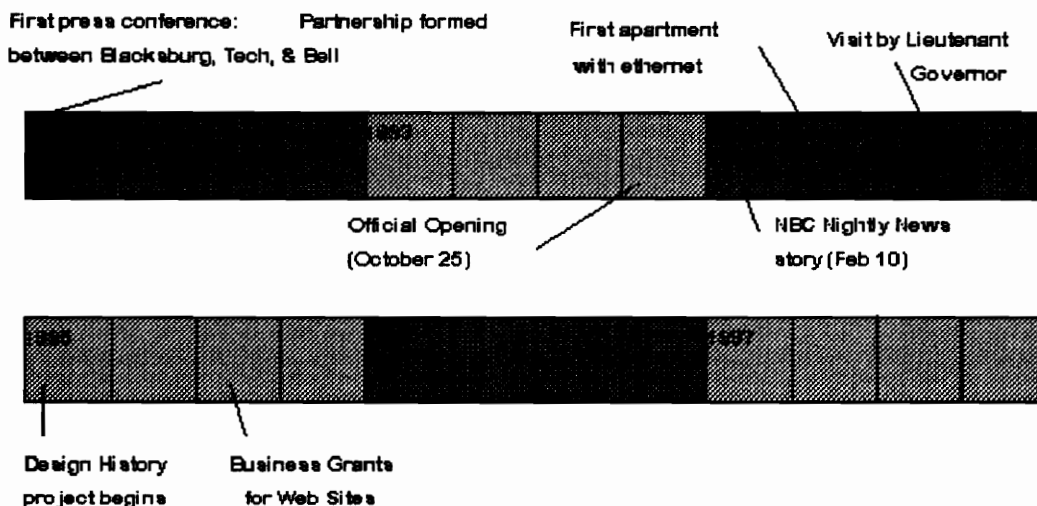
Click in a box to see a more detailed history for that quarter

Figure 5: Timeline view of BEV history

Other views of the HistoryBase include a Quarter-Year view, which is detailed view of the documents in a three-month period; a What's New view, which shows the recently contributed documents; and some BEV-specific views of the HistoryBase. These views of the system are discussed more thoroughly below.

Interface design and evaluation methodology

The design of the user interface was severely constrained by the current state of technology of the World-Wide Web. The language in which Web documents are authored, the Hypertext Markup Language (HTML), limits the number of interaction

techniques available. The interface design was based on conventional usability guidelines [Hartson93] as much as possible given the limitations of HTML.

The interface was evaluated as suggested by Nielsen [93] in his discussion of "discount usability engineering." Unlike formal usability engineering where many measurements are taken and statistical tests are carried out, discount methods rely on the designer to observe and interpret trial runs and to make the corresponding interface changes. This method was used because the goal of this project is to identify the major issues involved in designing a hypermedia history system, not to carefully assess the usability characteristics of a specific user interface. Thus the downside of not capturing every usability problem was judged to be offset by the time saved in doing a high level evaluation. Also the limited interaction techniques available using the Web limited the number of design choices so that in many cases there was only one way to implement a particular part of the system.

Interface walk-through

Any WWW browser can server as the interface to the HistoryBase (in this document the Netscape [Netscape96] browser is used for illustration). The basic element of the interface is a Web page; a collection of related Web pages is called a Web site. Figure 6 shows the first page a user sees when visiting the HistoryBase Web site. The title bar and menu bar at the top of this figure are parts of the Netscape browser, as well as the status section which covers the bottom quarter-inch of the screen. These parts of the interface

are not directly related to the HistoryBase and will not be discussed further. The Web

page itself starts with the HistoryBase graphic and extends down through the "Last

updated" line. Text that is underlined represents hyperlinks that will bring up another Web

page when clicked on.
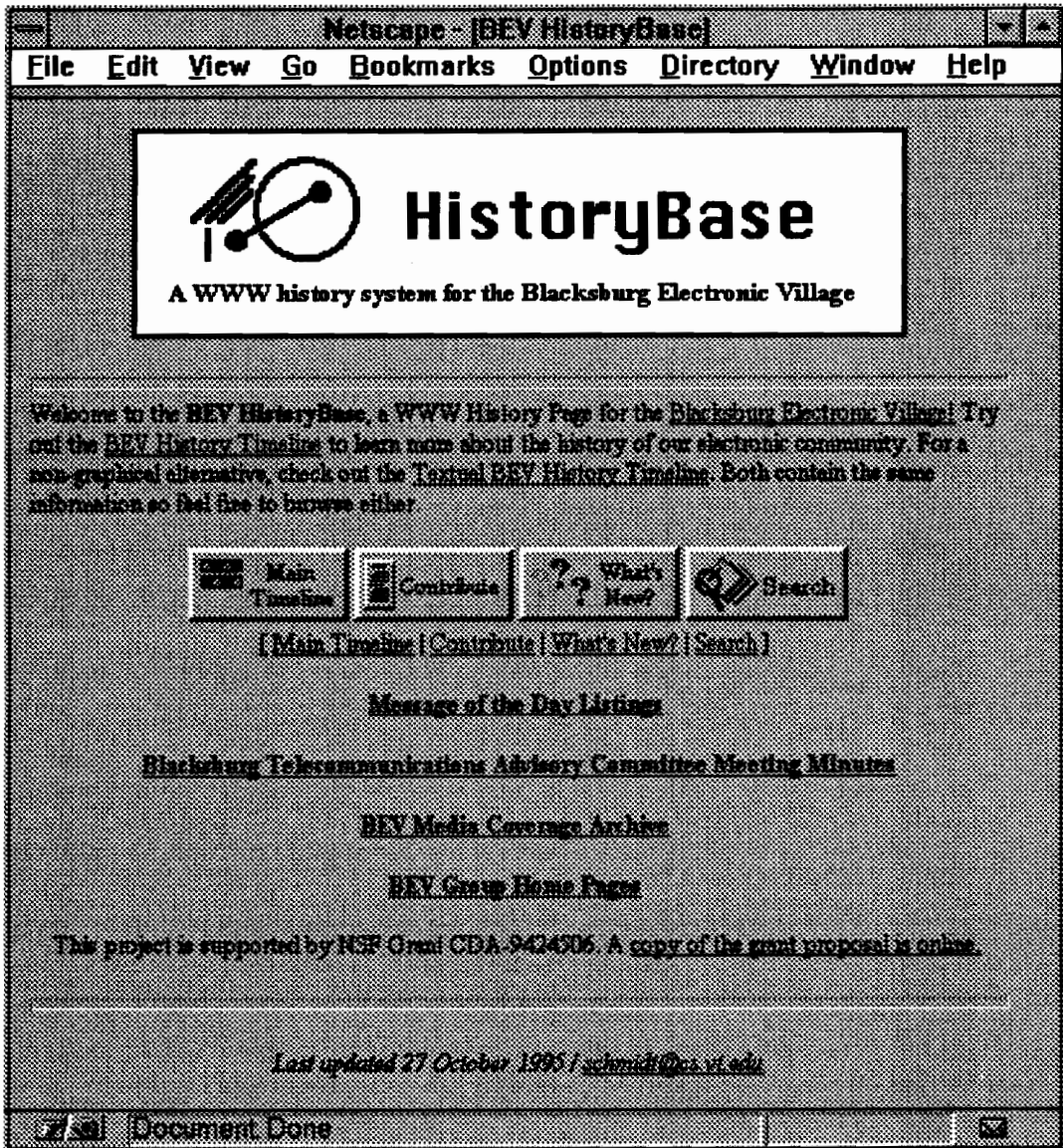
Figure 6: HistoryBase Main Page

HistoryBase Main Page

The HistoryBase Main page is the first page a user sees when visiting the HistoryBase

Web site. The BEV logo is used in the graphic at the top to serve a as a visual link

between the BEV and the BEV HistoryBase. The four buttons in the middle of the page, Main Timeline, Contribute, What's New?, and Search, are hyperlinks to the main parts of the system (textual links are provided below these buttons for those users whose Web browser doesn't support graphics). The four links starting with "Message of the Day Listings" and ending with "BEV Group Home Pages" are special sections of the HistoryBase that pertain specifically to the Blacksburg Electronic Village.

Main Timeline

The Main Timeline (see Figure 7) is a high-level chronological view of the HistoryBase. Key events are labeled and positioned on the timeline; a user can retrieve a list of all the documents for a particular quarter-year by clicking on the corresponding section of the timeline. At the top and bottom of this page, and every other page in the HistoryBase (except for the Main Page), is a navigational button-bar. This button-bar is placed at the top *and* bottom because it is possible for the upper button-bar to scroll off the top when the user moves through a long page.
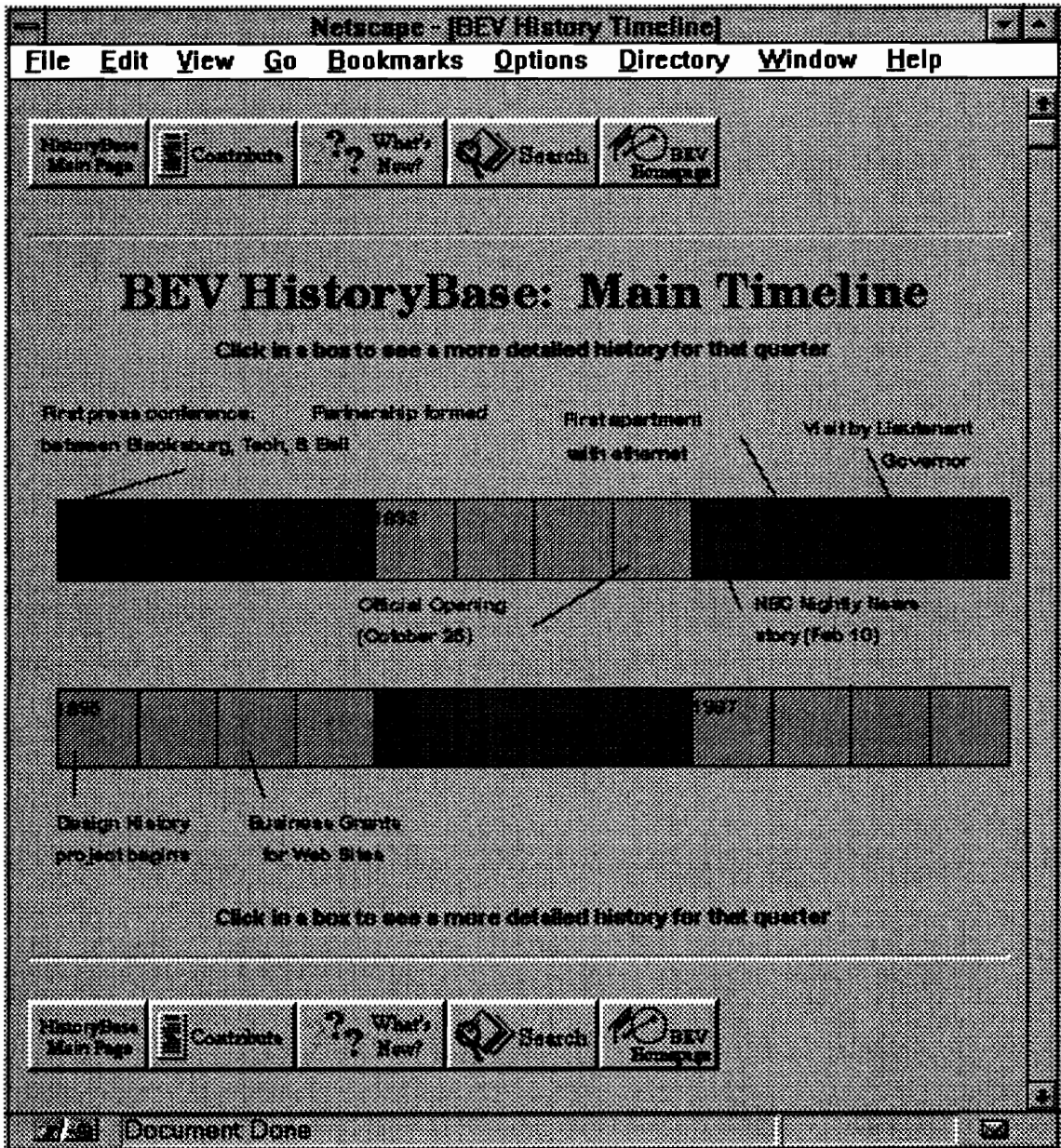
Figure 7: Main Timeline

Quarter-Year View Page

The Quarter-Year View Page is a document-level view of the HistoryBase for a three-month period. In Figure 8 the fourth quarter of 1995 is shown. At the top is the button-bar, and just below that is the "Quick-Click Timeline." The "Prior Quarter" and "Next Quarter" buttons are useful for moving backwards and forward through history; the Quick-Click Timeline allows the user to jump from one quarter-year view to another non-contiguous quarter with a single mouse-click.
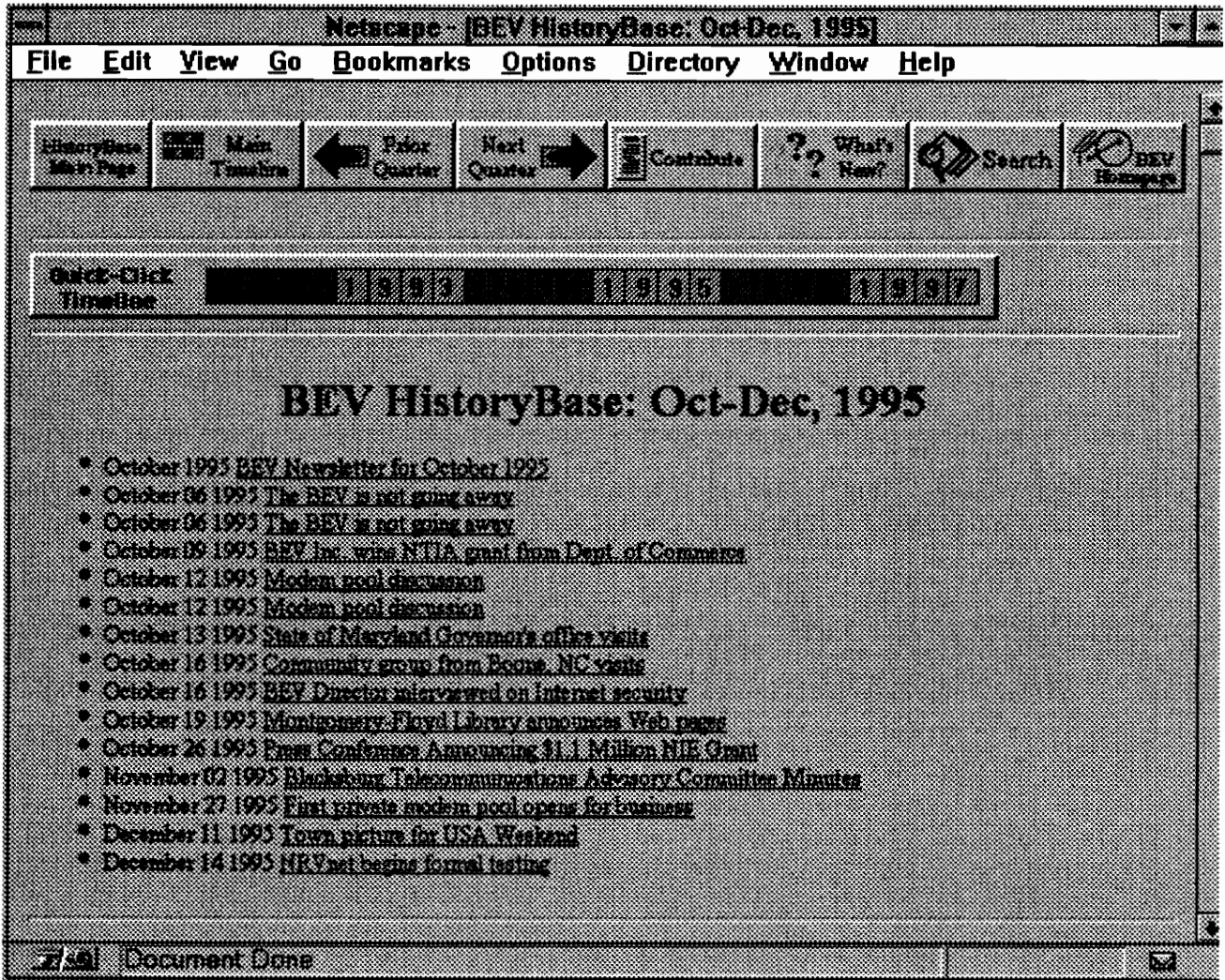
Figure 8: Quarter-Year View Page

Below the Quick-Click Timeline is the list of documents and events for this quarter-year.

On the left side are the document dates and to the right are the document titles, which are

also hyperlinks to the documents themselves.

Main Document Page

The Main Document Page is the display for a single document in the HistoryBase. A new button appears in the button-bar on this page, the Quarter Year View button. This is a hyperlink to the quarter-year view page in which this document resides. If the document is an event that spans over a quarter-year boundary, this button will link to the earlier of the two quarter-year views.

Below the button-bar is the meta-data for this document, followed by the body of the document. At the bottom is a link to the Annotation Page. The user can click on this link to make a comment about this document.

Figure 9: Main Document Page
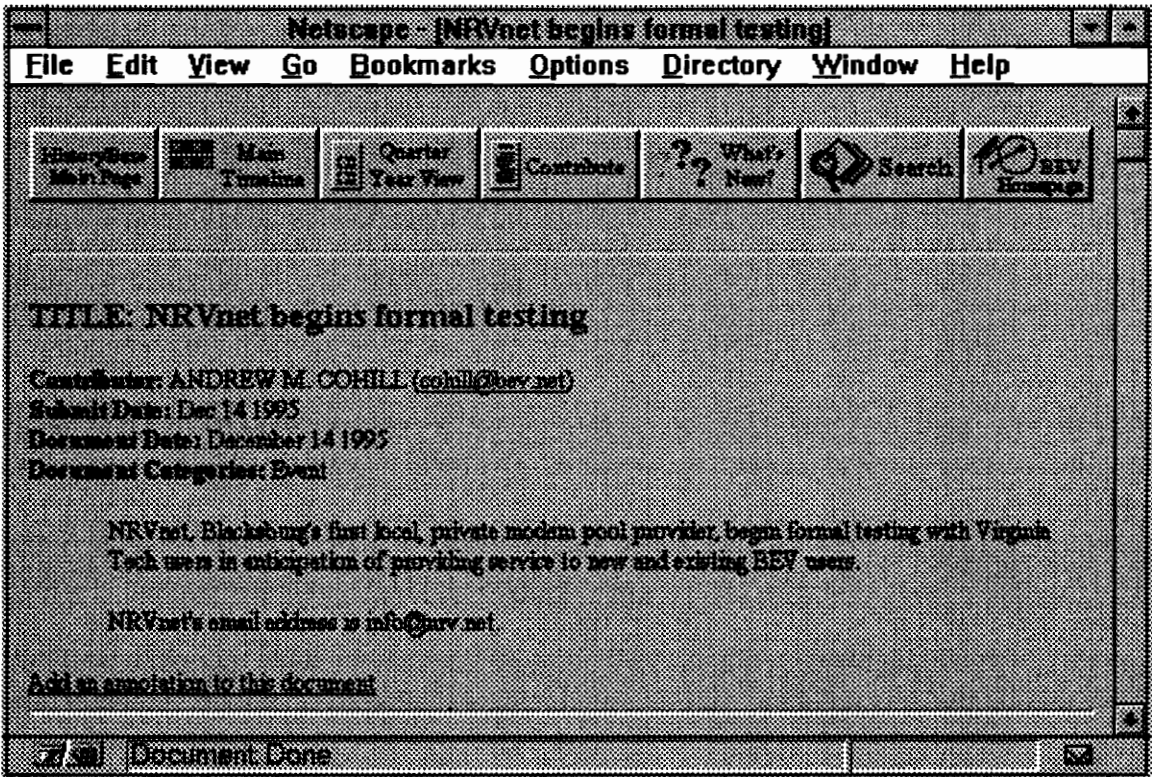
Annotation Page

The Annotation Page is used to make a comment on an existing HistoryBase document.

The body of the main document is displayed at the top (and any previous annotations), and

a text-entry box is at the bottom. When the comment is submitted, the main document is

re-displayed with the newest annotation at the bottom. Figure 10 shows an annotated

main document.

25
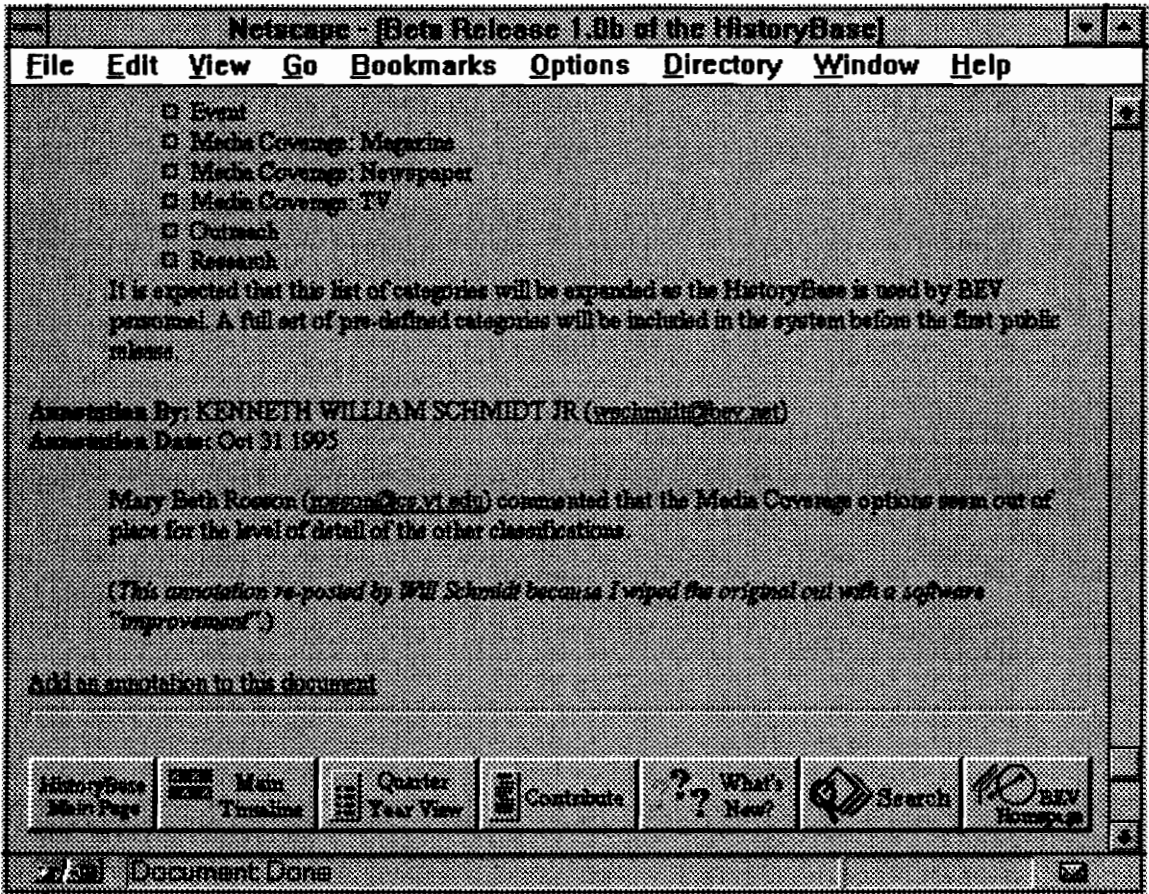
Figure 10: Annotated Main Document

Contribution Page

The Contribution Page (see Figure 11) is used to contribute a new main document to the

HistoryBase. At the top of this page is the Contributor Registration section. Anyone who

contributes to the HistoryBase must enter a PID (Personal Identification) and password

here. The system uses this information for authorization and to get the Longname (see

Figure 2) from the BEV authorization server.

Figure 11: Contribution Page

The next section is the Contribution Form. This is where the other Main Document

information (detailed in Figure 2) is entered. The document title and classification is

entered first, followed by the document or event date. If an event happens over a period of time then the ending date fields can be used to specify a time range. The body of the document can be entered in the text box at the bottom (in Figure 8 part of this text box has scrolled off the screen).

What's New Page

The What's New page is another time-oriented view of the document space. On this page documents are ordered by the date they were contributed, as opposed to the Quarter-Year view where documents are ordered by the document or event date. Clicking on the What's New button will always take the user to a list of the most recent contributions, whether these latest documents are clippings from a three year old newspaper or for an electronic town meeting held the evening before. The What's New page may be used by frequent HistoryBase users who just want to see a list of the latest documents available.
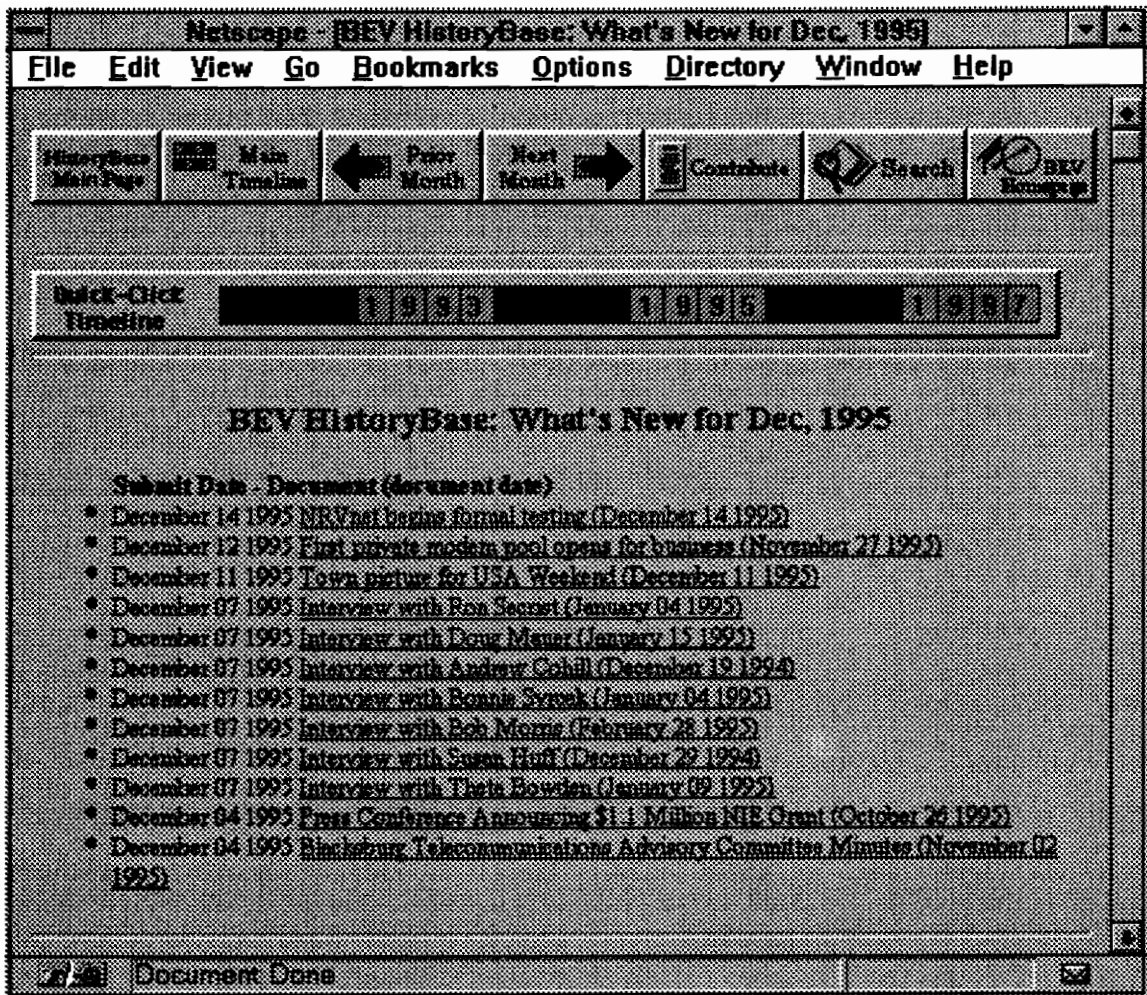
Figure 12: What's New Page

Search Page

The Search Page is used to find documents in the HistoryBase. Currently searches can be

performed on document titles, keywords (classifications), or in the body of a document.

Results of searches are displayed in ranked order by the raw number of occurrences of the

search terms in the documents.
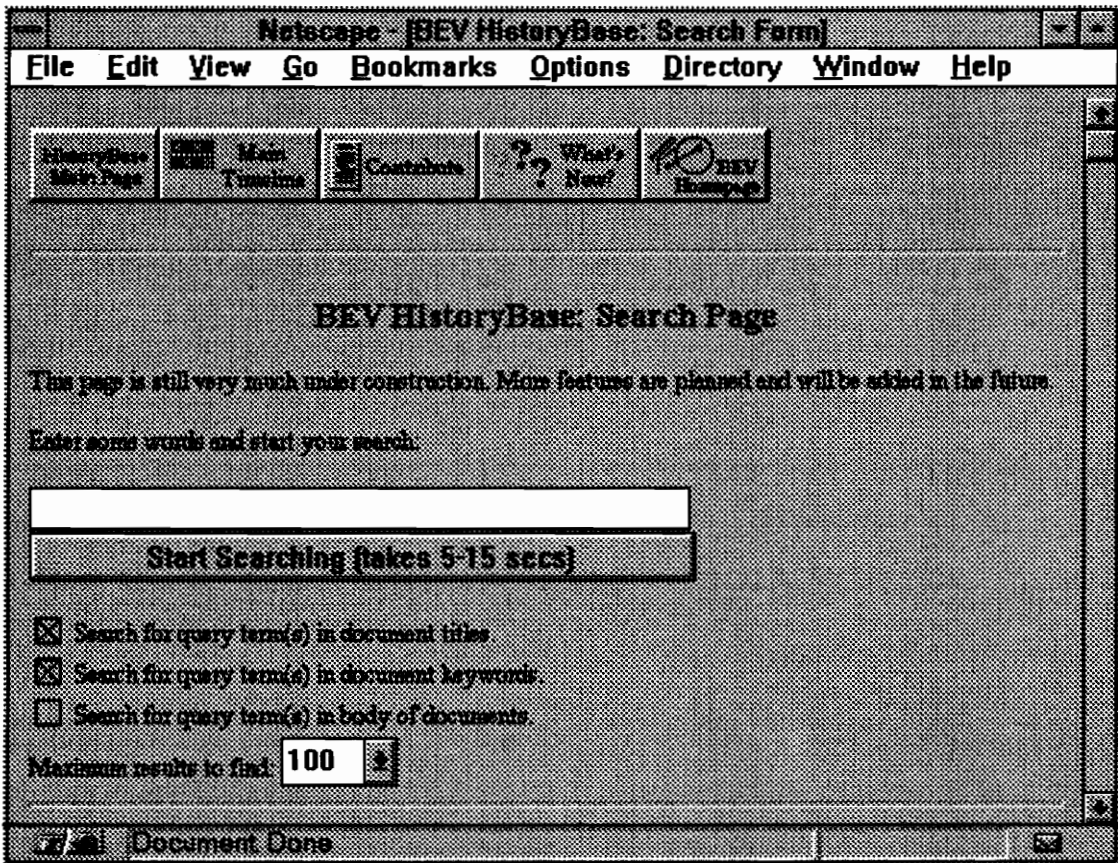
Figure 13: Search Page

Formative User Testing

The HistoryBase interface was tested informally on three subjects, two BEV project

personnel and one novice computer user. Many minor usability problems were found

during this testing, mainly in the placement of navigational buttons. One major problem

discovered was the need for a preview screen before the actual submission of a

contribution. Since a document can't be edited after it is contributed, users wanted one

last chance to look at the document before final submission.

Implementation

The HistoryBase was developed on a DEC Alpha, running the NCSA World-Wide Web server. The programs that make up the HistoryBase are written in Perl, and are accessed as Common Gateway Interface scripts through the Web server. Perl was used because its text processing features are more powerful and convenient than those provided by UNIX shell scripts or C. The back-end storage system is the Postgres database management system and the pgperl extension to Perl is used for interfacing between Perl and Postgres. All of the Web pages in the HistoryBase are generated by Perl scripts, usually by generating a query to Postgres and then formatting an HTML document.

Platform

The HistoryBase was developed on a DEC Alpha running the OSF/1 v2.1 operating system, but should be easily portable to any UNIX-based platform. In cases where system calls were necessary, care was taken to use only commands that are available on most, if not all, UNIX systems. The rare cases where system-specific procedures are needed have been documented and suggestions for how to port to other platforms included. For example user authentication in the HistoryBase is performed by calling the BEV authentication server. This server can be replaced by any program that is capable of responding to the same input patterns.

31

Database back-end

The HistoryBase uses the Postgres v4.2 database management system (DBMS) for raw data storage and query facilities. This choice was originally made because of the robustness of this system and because it supports some features of Object-Oriented (O-O) DBMSs; however the O-O capabilities of Postgres were never used.

All the HistoryBase programs that interface with Postgres were modularized so that a different data storage back-end can easily be plugged in. This was accomplished programatically by creating separate modules for DBMS interfacing, and also organizationally by storing the source code for DBMS-specific modules in a separate directory from the other HistoryBase code.

Programming language

The Perl programming is used because of its powerful text processing features. This text processing capability is used to convert raw query results into well-formatted HTML documents. The pgperl extension to Perl is used for interfacing with Postgres and handles the low-level communications between the two. Each Web page in the HistoryBase is generated "on the fly" by querying Postgres, formatting a Web page by adding HTML codes to the results of the query, and sending this document to the Web browser.

Common Gateway Interface and Web server

The Common Gateway Interface (CGI) is a standard for communication between Web servers and programs that are external to the Web server (usually referred to as CGI scripts). This interface allows the HistoryBase to use any CGI-compliant Web server; the original development was on the NCSA httpd v1.5 server.

Populating The Database

After the HistoryBase was up and running, several hundred documents were archived in the system by BEV project and HistoryBase personnel. These documents ranged from informal communications between BEV personnel to formal BEV documents like the vision statement and user guides. Table 1 lists the number and types of documents in the HistoryBase based on the keywords in each document. Note that the total number of press coverage documents is less than the sum of newspaper, magazine, and television. This is because a few of the press coverage documents were given more than one classification.

Table 1: Types of Documents in the HistoryBase

Number of documents in the HistoryBase	255
Press Coverage	119
Newspaper	86
Magazine	31
Television	6
Document	88
Events	33
Research	22
Outreach	14
BTAC Meeting Minutes	33

Two types of documents that made up a large percentage of this initial population are summaries of press coverage and the minutes of the Blacskburg Telecommunications Advisory Committee.

Press Coverage

From the first press conference where the Blacksburg Electronic village was announced (1992) until today, over a hundred news articles have been written about the BEV. A member of the BEV project team kept track of these press articles by making copies of the articles for the BEV archives and creating a flat-text file as an index. One of the first things that was put into the HistoryBase was a summary of each of these articles (and the full-text of the article when we had permission from the news source).

BTAC Meeting Minutes

The Blacksburg Telecommunications Advisory Committee (BTAC) has been involved with the development of the BEV since the early stages of the project. Many of the major events that shaped the project and the issues they raised were discussed by the BTAC. All of the meeting minutes from 1992 (the beginning of the BEV project) until the present were archived in the HistoryBase.

Usage

Figure 14 graphs the daily usage of the HistoryBase for the first 48 days of 1996. Note that this graph does not give a count of the number of people who accessed the HistoryBase. Each column represents the number of times any *page* in the HistoryBase was accessed on that day - one person could access only a few or many pages on a particular day.
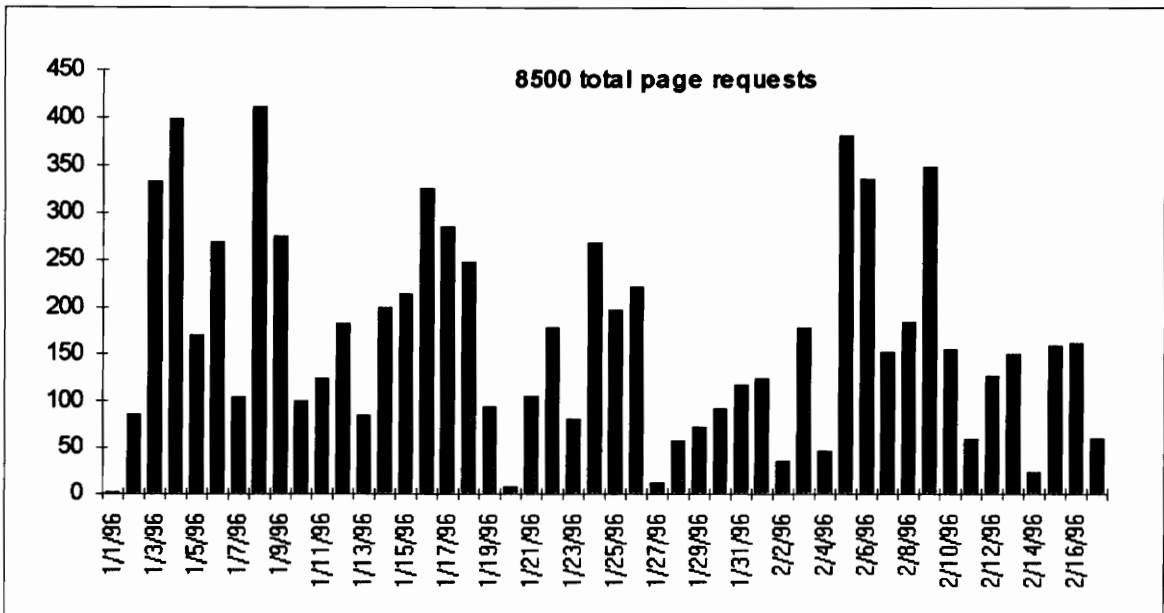
Figure 14: HistoryBase Activity (1/1/96 - 2/17/96)

Figure 15 shows the top ten requesting domains for the same 48 days. This breaks down the page accesses by the region from which the request originated. Each domain has two graph lines -- the longer graph line whose the total number of page accesses, and the shorter one shows the number of individual computers from within that domain that accessed the site. While the number of computers gives a different perspective on HistoryBase usage, it still does not give an accurate count of the number of people who used the system. In some cases many people may share a set of computers (e.g., a university computer lab) or everyone in a company may link from their desks through a common computer to gain access to the Internet. Still in other cases people from home gain access to the Internet by dialing to a modem pool where they may be serviced by a different computer each time they connect.
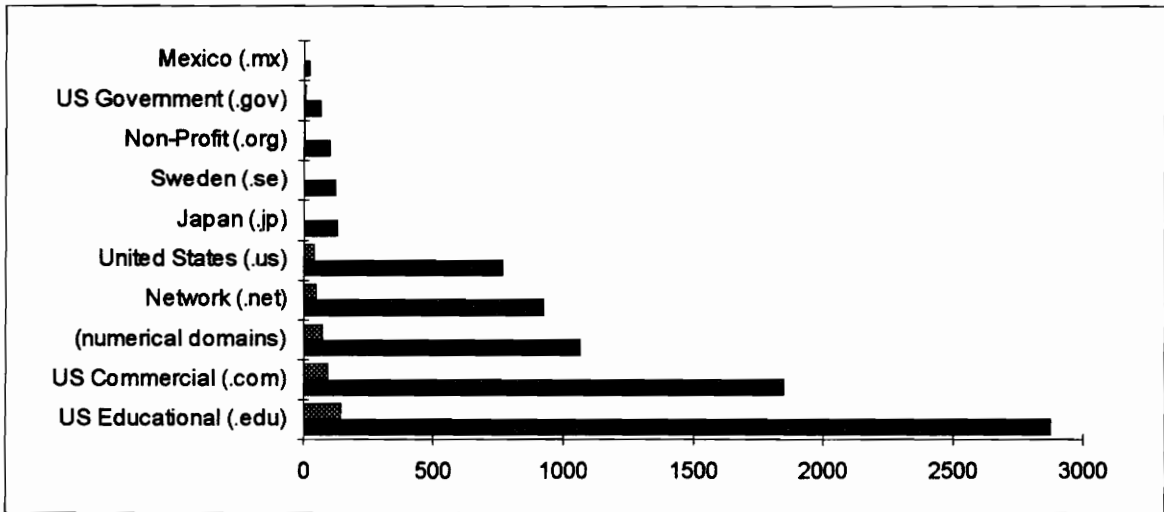
Figure 15: HistoryBase Domain Statistics (1/1/96 - 2/17/96)

Of the 2850 requests from the Educational domain, 2350 of these were from Virginia

Tech, which is a part of the BEV. Of the 800 requests from the Network domain, there

were 196 and 63 form the nrv.net and bev.net sub-domains respectively. These two sub-

domains represent the two main modem dial-ins for members of the BEV.

Issues

Many issues arose during the development of the HistoryBase. A discussion of a few of

these issues follows.

Use of the World-Wide Web

The World-Wide Web is being used as the implementation platform because this is the

least costly method to give HistoryBase access to the most people. A variety Web

browsers are freely available and many of the potential users of the system already use the Web. Also many of these same people view using the Web as something fun and interesting to do. It is hoped that using the HistoryBase will also be viewed as a fun activity and that people will enjoy making contributions to the system.

Pros and Cons of Anonymous Contributions

The HistoryBase does not allow anonymous contributions. There were several factors that effected this decision, including the ability to determine the validity and assign a value to a document, and the power if user identification in deterring system misuse.

If the HistoryBase allowed anonymous contributions there would be no way to judge the validity of a document. For example if the director of the BEV contributes a document on BEV personnel interactions, we will probably attribute more validity to that document than if the same document was contributed by someone who is not directly connected with the BEV personnel. The same is true for assigning a value to a document. Knowing who contributed a document helps us to assign a value to that contribution.

Anonymous contributions could also lead to a degradation of the quality of the contents of the system because there will be no person connected to some documents the system contains. This lack of connection could open the system to misuse. A prankster may find it funny to post pornographic material, or a disgruntled community member may post an obscene letter to the mayor.

However, by not allowing anonymous contributions, opinions on certain issues may not be recorded. For example, an employee may have a completely different view of an event or policy than his/her employer states publicly. This employee's job may be endangered by stating a contradictory view publicly. If this information is not recorded, the database will not hold a complete picture of all the forces that influenced the history. Of course, it would be impossible to record all possible views of an event, but the lack of anonymity may cause us to miss some of the important views.

Contributor Authentication

The HistoryBase authenticates contributors through a username and password scheme. This solution was chosen because all members of the BEV community already have a username and password they use when accessing the BEV and the Internet. The HistoryBase makes calls to the BEV authentication server to validate usernames and passwords.

Several potential solutions were considered before deciding to use the username/password scheme. One possibility is to send the contributor an email message whenever s/he makes a contribution. If a user receives an email message concerning a contribution made by someone else in his or her name, further action can be taken. This solution could cause problems for the HistoryBase system administrator. A single prankster could possibly submit tens or hundreds of documents on behalf of others. The HistoryBase system

administrator would need to remove all of these false submissions. This could also cause users to lose faith in the system.

Another authentication possibility is using the Internet address to identify a user. This is not possible with current Web browser technology so it is not an option for the HistoryBase, but it could be an option for future systems.

A third method for user authentication is to use digital signatures. A digital signature is a unique identifier that people can use for digital transactions. This method is the most attractive because it can be transparent to the user; however this technology is not yet readily available. Another possible drawback of this approach is the export laws for encryption technology used for digital signatures. If the history system will be an international one this type of authentication may not be legally possible.

The HistoryBase uses the username/password method for authentication, but there are drawbacks to this method. One is that passwords can be snooped and/or cracked, thereby defeating the security of a username and password. Another is that people tend to forget their passwords; password maintenance could become a burden on the HistoryBase system administrator. Finally, passwords can make the system seem less approachable to the lay-user. This could be the most damaging drawback to the HistoryBase because this would tend to discourage contributions. These issues are not as important because BEV users already have a username and password and the HistoryBase can tap into the BEV

user authentication system. The username/password scheme may need to be considered more carefully in systems where this is not possible.

Future Work

The HistoryBase is currently functioning as a part of the BEV Web site, but there are some improvements that can be made. Some of these are suggested improvements to various aspects of the system while others suggest functionality that can be added to make the HistoryBase more flexible.

Upgrading the Postgres DBMS

The Postgres DBMS has a limitation in that no record can contain more than 8 kilobytes of data. Since many documents are greater than 8K, all HistoryBase documents are stored in files outside of Postgres. Only the document meta-data is stored in the DBMS, along with a pointer (file name) to the body of the document.

A side-effect of this storage problem was an inability to use the searching powers of Postgres on the body of documents. Full-text searches are currently implemented using the UNIX *grep* facility. The limitations of this approach include poor scalability and slow execution times. The *grep* command is not readily scaleable because it does not use any type of index to perform a search. Every time the documents are searched with *grep* all the text is scanned. Thus the execution time of full-text searches increases linearly with

the number of documents. A more powerful search engine would work with document indices allowing searches to be performed more efficiently.

More Powerful Search Facilities

Another improvement to the HistoryBase could be the addition of fuzzy searching capabilities. The current HistoryBase searching mechanism depends on an exact match between query terms and the terms in documents. A search for the words "clubs" and "groups" will not return a document that contained references to "community organizations." Fuzzy searching would require application of artificial intelligence techniques to information retrieval to improve the chances of finding the documents the searcher really wants.

Support for Multimedia Contributions

Text documents are not the only type of history about the BEV that is being created. There are TV news stories, audio-visual presentations, photographs, and other multimedia aspects to the history of the BEV. These weren't fully implemented in the HistoryBase because of the limitations of the interaction techniques available in HTML -- text-based forms are supported, but audio and video file uploading is not (although Netscape just released a new version of its Web browser which has limited file uploading capabilities). Work was started on a way for multimedia objects to be submitted via email attachments,

but still needs work for full implementation. The addition of multimedia object contribution will be a significant step forward for the HistoryBase.

Better Integration of the Newsgroup Archives

Each night the HistoryBase automatically archives several key newsgroups pertaining to the BEV. At the moment all these news postings are just stored on the HistoryBase server -- there is no interface to browse and search these postings. The newsgroup archive program should insert the news postings into the HistoryBase so they can be accessed.

Extend the Automatic Archiving Capabilities

The HistoryBase has a few automatic archiving capabilities (like the newsgroup archiver), but there is a lot more information available that could be obtained this way. One possibility is to occasionally archive the top two levels of pages on the BEV Web Site. The HistoryBase already archives part of the top-level BEV Web page, where the "messages of the day" are listed. This is done every night, but the rest of the upper level pages could be archived on a less frequent basis.

Another idea is to give an easy method for internal BEV email messages to be archived in the HistoryBase. Perhaps BEV personnel could copy the HistoryBase on important email messages. In the future, automatic archiving will probably be the main method for populating the HistoryBase.

More Powerful Analysis Tools

The goal for the HistoryBase project was to build a tool for collecting, organizing, interpreting, and disseminating the documents and other materials pertaining the Blacksburg Electronic Village. A piece of this goal that has yet to be designed is an ability to analyze the contents the HistoryBase. A set of tools could be created that allow researchers to run complex queries on the data in the HistoryBase and to revise and refine these queries. These tools could also allow a person to leave behind the results of a query, and even the query itself, for others to view and interpret, and to use as the basis for further investigation.

Summary

This project has developed a historical database system that can be used for community networking projects and can be extended to other systems where recording historical data is of benefit. Although the HistoryBase is only starting to see use by members of the BEV community, we expect usage to rise with a little more promotion. The HistoryBase of the future have full multimedia features where users can contribute sound, images, and video clips. It will also have a full suite of analysis tools so the contents can more easily be explored and interpretations made. This project has taken the first big step towards an automated, self-documenting, hypermedia history system.

References

[Ackerman94] Ackerman, Mark S. (1994). Augmenting the Organizational Memory: A Field Study of Answer Garden, *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp. 243-252.

[Atwood95] Atwood, Michael E; Burns, Bart; Gairing, Dieter; Girgensohn, Andreas; Lee, Alison; Turner, Thea; Alteras-Webb, Sabina; Zimmerman, Beatrix (1995). Facilitating Communication in Software Development, Human-Computer Interaction Consortium (Snow Mountain Ranch, Frasier, CO, February 15-19, 1995).

[BEV96] Blacksburg Electronic Village (1996). World-Wide Web site. http://www.bev.net.

[Boulder96] Boulder Community Network (1996). World-Wide Web site. http://bcn.boulder.co.us/.

[Carroll94] Carroll, John M.; Cohill, Andrew Michael; Downey, Gary Lee; Fox, Edward A.; Rosson, Mary Beth (1994). NSF Grant Proposal: Building a History of the Blacksburg Electronic Village, 1994.

[Carroll95] Carroll, John M.; Cohill, Andrew M.; Rosson, Mary Beth; Schorger, John R (1995). Building a History of the Blacksburg Electronic Village, *ACM Symposium on Designing Interactive Systems*, 1995.

[CityLink96] The USA CityLink Project (1996). World-Wide Web site. http://banzai.neosoft.com/citylink/.

[Cutosky94] Cutkosky, Mark R.; Leifer, Larry J (1994). SHARE: A Scalable Framework and Methodology for Concurrent Engineering. http://cdr.stanford.edu/html/ONR/EndOfYearSummary94.html.

[Hartson93] Hartson, H. Rex; Hix, Deborah (1993). *Developing User Interfaces*. Wiley.

[LaLiberte96] LaLiberte, Daniel (1996). WWW Collaboration Projects. http://union.ncsa.uiuc.edu:80/HyperNews/get/www/collaboration.html.

[Morino96] Morino Institue Directory of Public Access Networks (1996). World-Wide Web page. http://www.cais.com/morino/htdocs/pandintr.htm.

[Nielsen93] Nielsen, Jakob (1993). *Usability Engineering*. Academic Press, Boston, MA.

[Netscape96] Netscape Communications Corporation (1996). World-Wide Web site. http://home.netscape.com/.

[Terveen95] Terveen, Loren G.; Selfridge, Peter G.; and Long, M. David (1995). Living Design Memory: Framework, Implementation, Lessons Learned, *Human-Computer Interaction*, volume 10, number 1, pp. 1-37.

[Vincent74] Vincent, John Martin (1994). *Historical Research: An Outline of Theory and Practice*. Lenox Hill Publishing & Distribution Company, New York, NY.

[Verzi92] Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors*, 34, 457-468.

[Webster94] Merriam Webster's Collegiate Dictionary, Tenth Edition, 1994.

[Weiss83] Weiss, H. Edmund (1983). Why we Still have so Little Technical Documentation. *Infosystems*, 30/5, 88-89.

Vita

Kenneth W. Schmidt, Jr. was born in Anderson, SC on March 13, 1969. He spent his formative years in Spartanburg, SC, where he graduated form Boiling Springs High School in June, 1987. He then attended Clemson University, where he participated in the cooperative education program and made the President's List three times. Between his December, 1992, graduation from Clemson and enrolling in Virginia Tech in January, 1994, Schmidt worked as a Computer Specialist for the National Climatic Data Center in Asheville, NC. After graduating from Virginia Tech in May, 1996, he plans to continue growing the company he founded in during the summer of 1995.