

**IMPROVING GRADUATE EDUCATION THROUGH DIGITAL LIBRARY TOOLS**

**by**

**Todd Hill Miller**

**Thesis submitted to the Faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of  
HONORS BACCALAUREATE  
in  
Computer Science and Computer Engineering**

**APPROVED:**

---

**E. A. Fox, Chairman**

**May, 1999**

**Blacksburg, Virginia**

**Key Words: Annotation, Electronic theses and dissertations, Evaluation**

**Copyright 1999. Todd Hill Miller**

**Abstract**

**IMPROVING GRADUATE EDUCATION THROUGH DIGITAL LIBRARY TOOLS**

**by**

**Todd Hill Miller**

**E.A. Fox, Chairman**

**Computer Science and Computer Engineering**

**(ABSTRACT)**

As technology advances and the World-Wide Web grows in popularity, more and more research can be conducted on-line. This is being made possible through the creation of digital libraries, such as the Networked Digital Library of Theses and Dissertations (NDLTD). NDLTD provides Web access to a collection of thousands of electronic theses and dissertations (ETDs) catalogued and stored in an electronic format. In theory, the increased access to these works will increase the quality of graduate education and scholarly exchange.

Use of these materials can be encouraged through the development of digital library tools. These tools are able to provide graduate students with capabilities not easily achieved with paper theses and dissertations. Examples of possible tools are ones that can find the most influential articles for a group of documents or identify open problems. This study examines the benefits and usability of one digital library tool in particular, the ETD Annotation System.

The ETD Annotation System was developed to enable users of NDLTD to annotate documents in the collection. Annotations can be public, shared with a group to promote discussion and scholarly exchange, or they can be private to keep track of ideas or thoughts stimulated by a document. This study enlisted a group of graduate student volunteers to use the system in a practical manner and give feedback on their experiences. The objective was to identify perceived benefits from an annotation system as well as what might enhance use of the collection.

Results of the study show that graduate students found the annotation system to be both beneficial and easy to use. They commented that it would help address one of their biggest problems, keeping track of the masses of information that they go through in the course of their research. They also realized the potential for discussion and interesting perspectives to arise out of the deployment of an annotation system. It excited them that they could come back in the future to see what other people thought of it, or to learn how it had been used in the research of others.

Thus the study showed that digital library tools can bring added benefits to NDLTD, and encourage further use. While it was not possible to prove long term increases in the quality of graduate education, this study provides some evidence that graduate students may work more effectively through the use of digital libraries.

### **Dedication**

To Mom and Dad

For all the love and support you've given me.

### **Acknowledgements**

This work would not have been possible without the assistance of many people. I'd like to thank Dr. Edward Fox most of all for all of his extraordinary work, late night copy editing, and patience with me over the past three semesters. I'd also like to thank Gail McMillan, Dr. John Eaton, Neill Kipp, Constantinos Phanouriou, Ohm Sornil, and Paul Mather for all of their assistance with my research.

## Table of Contents

<b>1</b>	<b>INTRODUCTION.....</b>	<b>1</b>
1.1	PROBLEM STATEMENT .....	2
1.2	PURPOSE AND OBJECTIVES .....	3
1.3	SCOPE .....	3
1.4	STUDY OVERVIEW .....	3
<b>2</b>	<b>REVIEW OF LITERATURE .....</b>	<b>5</b>
2.1	DIGITAL LIBRARIES .....	5
2.2	AREAS RELATED TO POSSIBLE DIGITAL LIBRARY TOOLS .....	10
2.2.1	<i>Agents.....</i>	<i>11</i>
2.2.2	<i>Collaboration Over the Internet.....</i>	<i>13</i>
2.2.3	<i>Summary of Areas Related to Possible Digital Library Tools .....</i>	<i>15</i>
2.3	ANNOTATIONS .....	16
2.3.1	<i>Paper-Based Annotations.....</i>	<i>17</i>
2.3.2	<i>Application-Based Annotations.....</i>	<i>17</i>
2.3.3	<i>Internet-Based Annotations.....</i>	<i>19</i>
2.3.4	<i>Summary of Annotation Literature.....</i>	<i>20</i>
<b>3</b>	<b>PRELIMINARY STUDY .....</b>	<b>22</b>
3.1	INTRODUCTION .....	22
3.2	PROCEDURE .....	22
3.3	RESULTS .....	23
3.4	CONCLUSIONS.....	23
<b>4</b>	<b>THE ANNOTATION SYSTEM .....</b>	<b>25</b>
4.1	INTRODUCTION .....	25
4.2	DESIGN CONSIDERATIONS .....	25

4.3	IMPLEMENTATION OVERVIEW.....	26
4.4	FUTURE WORK .....	29
<b>5</b>	<b>SUMMATIVE EVALUATION .....</b>	<b>31</b>
5.1	SAMPLE SELECTION .....	31
5.2	INSTRUMENTATION .....	31
5.3	DATA COLLECTION PROCEDURES.....	32
5.4	LIMITATIONS.....	33
<b>6</b>	<b>RESULTS .....</b>	<b>34</b>
6.1	DESCRIPTION OF PARTICIPANTS.....	34
6.2	SUMMARY OF RESPONSES.....	34
<b>7</b>	<b>DISCUSSION AND IMPLICATIONS.....</b>	<b>36</b>
7.1	RESPONDING TO RESEARCH QUESTIONS .....	36
7.2	IMPLICATIONS OF THE STUDY .....	37
7.2.1	<i>Implications for Future Annotation Systems .....</i>	<i>37</i>
7.2.2	<i>Implications for Future Digital Library Tools .....</i>	<i>37</i>
<b>8</b>	<b>REFERENCES.....</b>	<b>39</b>
<b>9</b>	<b>APPENDIX A – ANNOTATION SYSTEM IMPLEMENTATION .....</b>	<b>40</b>
<b>10</b>	<b>APPENDIX B – USABILITY STUDY QUESTIONNAIRE .....</b>	<b>41</b>
<b>11</b>	<b>APPENDIX C – SCENARIOS FOR PRELIMINARY STUDY .....</b>	<b>45</b>
<b>12</b>	<b>VITA.....</b>	<b>48</b>

# 1 Introduction

A vast universe of knowledge is contained in the numerous theses and dissertations produced each year at colleges and universities throughout the world. In the past, these works typically sat hardly touched in their respective libraries. Companies like UMI were started that would microfilm and sell copies of dissertations, which improved the amount of use that these documents received. Today, with the Internet and computer technology, it is possible to store theses and dissertations electronically and provide the world with quick access to this wealth of knowledge. The Networked Digital Library of Theses and Dissertations (NDLTD) is an initiative seeking to help universities improve education through digitizing their graduate research.

One of the focuses of the NDLTD project has been on examining the new possibilities of utilizing this information, given its electronic form. The idea is that graduate students could conduct higher quality research through use of digital libraries of electronic theses and dissertations (ETDs). New tools that take advantage of the digital media therefore should further increase the benefit that graduate students gain. An example of such a tool is an annotation system that allows users to write annotations or summaries on ETDs that they've browsed in an online collection.

Research on digital documents in general has proceeded since the 1950's. For example, many tools already support writing, indexing, and annotating. However, much of this research has yet to be applied to a digital library that is accessible over the World-Wide Web. As the Web grew in popularity and size, demand increased for tools that extend its value. Many tools had been developed to enhance users' experiences on the Web, such as advanced search engines and personalized agents [Lieberman, 1997]. Some research has been conducted in developing systems for annotating Web pages, but many require a special Web browser to function [Roscheisen, Winograd, Paepcke, 1995]. Most of the tools thus far have addressed enhancing the experience of the entire Web, rather than for a particular collection of documents online.

This study has focused on supporting graduate students in their educational research through the development of an annotation system for NDLTD. Annotation methods were reviewed to help determine how to design a tool specifically for a collection of ETDs. The tool required no special Web browser, promoted collaborative research, and was designed with graduate students in mind.

## 1.1 Problem Statement

The problem addressed is how to best support a graduate student's educational research. The approach taken is to encourage graduate students to build and use a collection of electronic theses and dissertations, such as that of NDLTD. The use of NDLTD is further encouraged by providing additional benefits that are not as readily realized by the paper counterpart of the collection. The problem of how to best support research with NDLTD and provide these extra benefits is solved by developing digital library tools designed for graduate students. This approach involves two parts, each of which has a set of questions to be answered:

- 1) Planning the tools to support research with a digital collection of ETDs.
  - a) What tools would students find most useful?
  - b) What tools would be the most feasible to implement with today's technology?
  - c) How can these tools best assist students in their research?
- 2) Developing one of the more important tools.
  - a) Which tool should be developed first?
  - b) How is the information generated by these tools handled?
  - c) What are the benefits gained by graduate students through use of these tools?

This study is important for 3 reasons. First, it demonstrates a method of determining how to go about designing tools that will assist graduate students conducting research with a library of ETDs. No previous research was found that focused on assisting graduate students in using a digital library. Secondly, this study will show the potential for using annotations as a means of collaboration within NDLTD. Annotations are a proven hypertext method to encourage collaboration, but no one has looked at what would occur within a digital library of ETDs. Lastly, it will provide some evidence as to the benefits graduate students receive by using NDLTD. It is an assumption that benefits would be gained, but very little has looked at what those might be or whether they exist.

## 1.2 Purpose and Objectives

The purpose of this part of the study was to find ways to take advantage of a digital library of electronic theses and dissertations through the development of an annotation system. This focus was based on a preliminary study of the needs of graduate students and existing technology to find the areas that could be most readily built upon. The findings were presented to the steering committee of NDLTD at the Fall 1998 meeting, and it was their recommendation that the annotation system be implemented first. After the system was designed and tested, graduate students participated in a usability study to determine its benefits. The following objectives were examined in this study.

- 1) To demonstrate that a tool could be developed for annotating ETDs, with minimal burden on the users.
- 2) To demonstrate that an annotation system for a collection of ETDs is beneficial to graduate students as well as faculty members.
- 3) To demonstrate that with such additional benefits, using NDLTD to support research would be more attractive to graduate students.

## 1.3 Scope

The scope of the study has been limited in three ways. First, the annotation system was developed using the collection of ETDs available through NDLTD. Thus, it may not be applicable to other digital library collections. Secondly, the study has been restricted to persons at Virginia Tech, specifically graduate students and faculty members. Therefore, the results may not be generalizable to other populations. Lastly, only a small number of graduate students participated in this study. These students represent a select sub-population. It is possible that the results may have been different if a less select population had been used.

## 1.4 Study Overview

This report has six remaining parts. The second chapter is a review and analysis of supporting literature. The preliminary study and the design of the annotation system are presented in Chapters 3 and 4, respectively. The

fifth chapter outlines the evaluation study. Analysis and results are presented in Chapter 6. The final chapter is a summary of findings, discussion, and suggestions for future research.

## 2 Review of Literature

In order to explore how to support graduate-level research with an enhanced collection of ETDs, it was first necessary to examine the literature on what has already been done with digital libraries and tools for them. Three major groups of studies emerged in this review: studies about digital libraries, studies covering areas related to possible tools for a DL, and studies that examined annotation systems. Each of these areas will be presented, along with a summary of findings, and how the current study will extend beyond prior investigations.

### 2.1 Digital Libraries

A digital library is a library consisting of digital documents that are stored and accessed online. The basic concept of such a library has existed for quite some time, but has become increasingly popular with the growth of the Internet. There are several works that were reviewed in preparation for this study. They examined how to build and maintain digital libraries, as well as how to address some of the future problems that digital libraries will face.

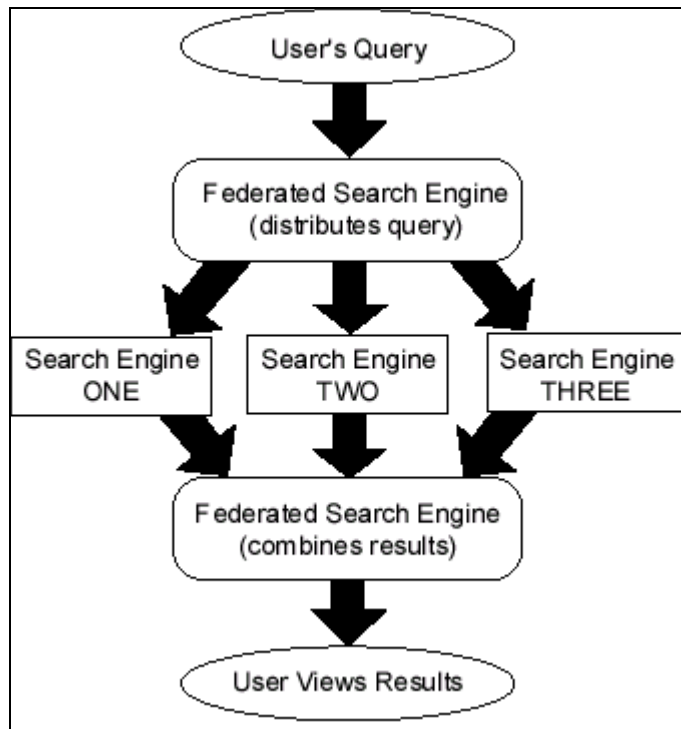
#### 2.1.1.1 Large-Scale Digital Libraries

One such paper is “Building Large-Scale Digital Libraries” by Schatz and Chen. This paper gave an excellent introduction to the US Digital Library Initiative (DLI). The DLI is related to government efforts to develop the National Information Infrastructure (NII) [Schatz and Chen, 1996]. The goal of the NII is to “bring the highways of knowledge to every American”, and the chosen method of delivery is through digital libraries. The first phase (1994-1998) of DLI involves several government agencies, such as NSF, DARPA, and NASA, as well as six universities. Each university was involved in building large collections as testbeds for further research into digital libraries. According to Schatz and Chen, there were two major areas of focus in their project.

The first area concentrates on how to develop new methods to index documents. Most digital documents are currently indexed based on their significant words, so a successful search depends on matching the user’s keywords with the document’s significant words. Traditional libraries have human indexers that can generate other words, called subject descriptors, that can improve a search by increasing the likelihood that the user’s keywords will match [Schatz and Chen, 1996].

The second area of research examines the process of federating the collections of digital objects. Federating is the term for a distributed search across several digital libraries in a manner that makes them appear as one large

library. In order for federated searches to be possible, a common gateway has to be developed that can translate a user's query into a query that the search engines of each individual collection can understand. An illustration of how a federated search works is provided in Figure 2-1. It shows how a user's query is taken and distributed to a set of distinct search engines, with the results combined back together to make it seem as if they were all part of one search [Schatz and Chen, 1996].



**Figure 2-1 Federated Search System**

These concepts of indexing and federating have expanded the amount of research being done on information retrieval. The basic principles in this field have gone largely unchanged over the last 30 years, but are now being forced to adapt to deal with the Web and large scale distributed digital libraries. However, the issue of information retrieval is not the only problem facing digital libraries.

#### 2.1.1.2 PURLs and Metadata

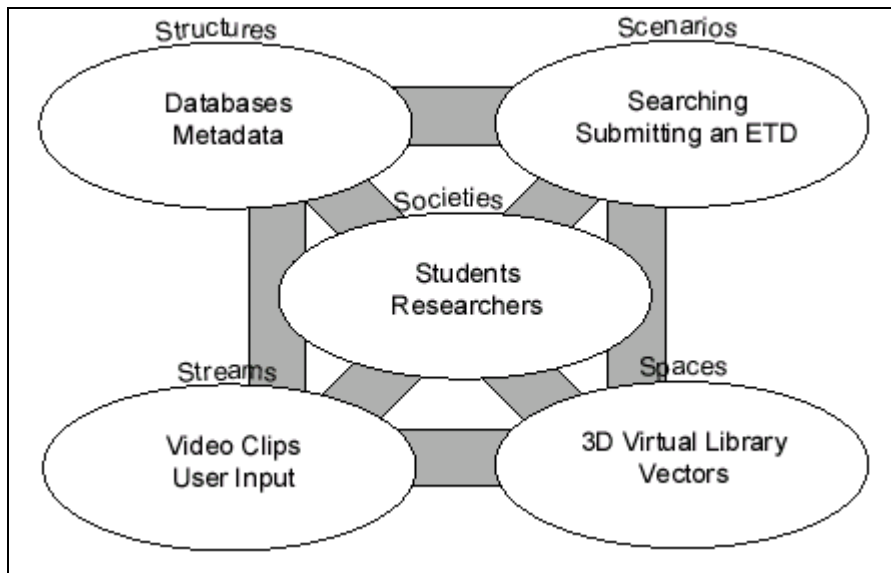
Other issues currently hinder progress towards the goal of a worldwide, searchable, information space, as shown in Miller's paper entitled "W3C and Digital Libraries" [Miller, 1996]. This paper highlights two major difficulties that currently face researchers, that of persistent names and the appropriate metadata for documents.

Persistent names are at the very core of digital libraries. A name is the means of access to an object, just as a URL enables accessing documents on the Web [Miller, 1996]. Documents in a digital library need to have a name that meets two conditions, that no other document will have the same name, and that the name will last forever. One current answer to this problem is the Persistent URL (PURL). The PURL is a unique name for a document, that will never change, even though the location of the document itself may. As long as the user knows the document's PURL, he will always be able to access it.

Miller also addresses the issue of metadata, which relates to the information retrieval issues previously addressed. Metadata is information about a document, such as the author or title [Miller, 1996]. However, deciding exactly what should be part of the metadata is a subject of great debate. Solutions are beginning to surface with the introduction of proposed standards such as the Dublin Core and the Warwick framework, which outline sets of metadata [Miller, 1996]. However, the problem is further complicated by requirements for encoding this metadata and then transmitting or receiving it. Miller suggests that the Platform for Internet Content Selection (PICS), be used for this task. PICS was originally developed to assign content ratings to Web material in order to protect children. However, the ratings themselves are in essence metadata, so PICS serves as a basis for the exchange of other metadata as well. The Resource Description Framework (RDF), developed by W3C, extends the earlier work on PICS. As we delve into the concepts of metadata, the need to further define the concept of a digital library becomes readily apparent.

#### 2.1.1.3 Models of Digital Libraries

One view on the theory of Digital Libraries is the 5S model, as outlined in a paper by Neill Kipp entitled "Flying Saucers, Wizards, and Deep Space". The 5 S's are as follows: streams, structures, spaces, societies, and scenarios. Figure 2-2 illustrates how the 5S's interact and relate to each other. Each of these arise from the requirements for a digital library to handle cataloging, scanned texts, images, multimedia objects, virtual environments, hypertexts, databases, and software.



**Figure 2-2 4S Model of Digital Libraries**

Kipp states that with the broad range of materials that have to be dealt with, the “book” metaphor for digital libraries is too restrictive, as there is the possibility for much more interaction than with a paper book. This interaction flows over the “streams” of the 4S model. It encompasses all of the information exchanged (e.g., over networks or cables) when utilizing a digital library, such as keystrokes and mouse gestures, as well as video and audio. How these streams are dealt with is an important consideration in the design of a digital collection.

The next part of the model, structure, is defined by Kipp as “anything that can be represented by an annotated tree or directed graph”. Within a digital library, many different structures may be used. Some structures will contain information about how the documents are organized, while another example would be a structure for the fields of metadata that are used for searching and indexing the documents. Hypertexts and database management systems depend on structures.

Spaces, such as vector, probability, and measure spaces make up the third component of the 4S model. Kipp states that such spaces are used to “facilitate the library organization, management, and presentation”. Notions of location or nearness depend upon the concept of space. Finally, the fourth part of the model is the scenarios. These are used to model how streams of data flow in and out of a digital library, seeing how they are always actively exchanging information and characterize the services of the library.

Kipp views the benefits of a model such as 4S as providing a basic groundwork for a digital library. The model can provide a way to formalize everything from user interface to evaluation, which is a key aspect in ensuring that the benefits to the user are maximized. Through evaluation, digital libraries can be further enhanced. Kipp suggests utilizing some of the basic tenets of usability engineering. Some of these principles are to consider the human first in design, be consistent, make all messages clear and concise, and provide the user with as much control as possible.

These are guidelines that were followed in the development of the annotation system for this study. Development of any system requires in depth thought and planning to ensure that the end result will be usable. This is called human factors engineering. These principles are currently being applied by groups such as the Network Digital Library of Theses and Dissertations (NDLTD), to build large collections of ETDs. The NDLTD initiative is outlined in the paper “National Digital Library of Theses and Dissertations: A Scalable and Sustainable Approach to Unlock University Resources”.

#### 2.1.1.4 NDLTD

The goal of NDLTD is to improve graduate education with a vast distributed digital library that consists mainly of theses and dissertations. They plan to accomplish this by not only making more information available to graduate students for conducting research, but also to educate them in how to publish documents electronically [Fox et al, 1996]. Fox states the vision is that students will be able to “find the full texts of related works easily, to read literature reviews prepared by their peers, and to follow hypertext links to relevant data and findings”. Among other expected benefits to graduate students are: the increase in citations to their ETD from other works, the notification through email of new studies related to their research, and lowering of costs of preparation and submission of their thesis or dissertation.

However, a project like NDLTD does require extensive change and support at the university level. Some infrastructure will have to be put into place before the efficient production and archiving of ETDs can take place. First, software such as Adobe Acrobat Exchange may be purchased in order to support the authoring of ETDs. Access to multimedia labs as well as technical assistance should be provided to handle the various types of multimedia objects, such as images and audio clips. In the end, universities can benefit greatly from NDLTD. They

can publish the research results of all graduate students at minimal cost, as well as free up valuable shelf space in their libraries.

Fox points out that there are many open areas for further research and development with NDLTD. These areas are best categorized into three groups: software, standards, and systems. Software is needed to deal with increasingly complex multimedia documents. Issues such as the size of documents and maintaining a digital library collection are areas within this. The need for new standards to be defined is also great. Without the proper standards in place, the content in a digital library could become virtually unusable as technology improves. Good standards will ensure that the content is preserved for many decades to come. Lastly, digital library systems are needed for the long term success of NDLTD. A computer system is needed that can handle large databases of various multimedia object types. In addition, such a system should ultimately be commercially supported in order to ensure continuing support and existence.

#### 2.1.1.5 Summary

In summary, there has already been a great deal of research done into digital libraries [Fox & Sornil, 1999]. It is obvious that there are many questions posed that have yet to be answered. Some of the best examples of these are how to preserve documents in a digital library over time and how to best index and search them. Most of the research reviewed thus far has focused on the library technology itself, basically duplicating in a digital manner what can be performed in a traditional library. This study will extend upon the work already done on digital libraries by pushing into an area not as extensively covered: the area of tools that enhance the digital library experience and that take advantage of the electronic media. The annotation system developed and analyzed in this study is just an example of one such tool.

## 2.2 Areas Related to Possible Digital Library Tools

Many other possible tools were explored in a preliminary study, before the decision was made to focus on annotations. These tools represent some of the enormous potential that exists regarding utilizing a digital library. Examples include agents that can find open problems, or define jargon, as well as new ways to present and communicate information, such as programs to compare and contrast two ETDs or an online public forum for

discussions. In researching these tools, the literature seemed to fall into two primary groups, agents and Internet-based collaboration.

### 2.2.1 Agents

Agents are a powerful tool for automating complex tasks. An agent is a computer program that works to perform a particular task for the user, in an attempt to save the user time. These programs can run with or without the direction of the user. One that runs with little or no guidance is termed an “autonomous” agent. These agents anticipate the needs of the user and try to provide information in a just-in-time fashion. An example of this type of agent is one that notifies the user when a regularly visited Web page has been updated. Other agents work with the user to accomplish a task faster, such as setting up a chart in a spreadsheet by guessing at the correct axes and labels. Table 2-1 illustrates the differences between the two types of agents.

**Table 2-1 Comparison of Types of Agents**

Characteristic	Autonomous Agents	Interface Agents
Visibility	See only its results	Watch it work
Level of Interaction	Low	High
Types of Tasks	Continuous	On demand

#### 2.2.1.1 Autonomous Interface Agents

Henry Lieberman wrote a paper entitled “Summary of Autonomous Interface Agents”, which provides a good overview of what agents are and how they can be applied to the Web in general. He suggests combining the aforementioned two types of agents, termed the autonomous interface agent [Lieberman, 1997]. This type of agent operates some part of the interface in an autonomous fashion, such that the user can see what it’s doing, while at the same time the agent can observe the user. Lieberman suggests that these agents are not suited for situations where their “decisions” are critical. Rather, these agents should be presenting a list of best guesses to the user in a timely fashion. The user must take the time to review the suggestions provided by the agent and decide whether or not to act upon them. This also raises the point that the agent should not be overly distracting as to take away from the task being performed by the user.

According to Lieberman, agents should be able to refine their guesses based on the user’s actions, rather than having the user “teach” the agent [Lieberman, 1997]. This is based on the premise that users want to minimize

the amount of “fooling around” that they have to do with the agent in order to get it to be useful. It is proposed that users will most likely only use an agent if it can be immediately useful to them.

Following his own advice, Lieberman constructed an autonomous interface agent that browses Web pages that are just a couple of links away from the one being viewed [Lieberman, 1997]. This agent, called Letizia, scans the pages for keyword content in relation to the current page, and presents a list of pages that it determines to be similar in a side window. This is just one example of how an agent can be used in the context of the Web, but they can be used in the environment of a digital library.

A tool employing agents in a digital library setting could prove to be extremely useful. Agents could find related documents in the collection, find the section of a cited document that the citation comes from, or even assist in the authoring and publication process.

#### 2.2.1.2 Agents Used in Digital Libraries

The University of Michigan Digital Library project is working towards creating a digital library structure based on autonomous agents. An overview of this project was published in the article “Toward Inquiry-Based Education Through Interacting Software Agents”. This paper presents the idea of a distributed agent architecture which is laid upon a digital library [Atkins et al, 1996]. The rationalization for such a thing is that technology changes so rapidly that it would be difficult to constantly update the software to accommodate new innovations. Atkins’ solution is to require the DL architecture to perform generic management operations, instead of adopting specific standards. These operations would be distributed to numerous highly-specialized agents. New services could be built upon these agents, without disturbing the overall system. The agents would have a generic communication interface through which services could communicate in order to have them accomplish their library task.

Atkins divides these digital library agents into three groups: user interface agents, mediator agents, and collection interface agents. The User Interface Agents (UIA) control how information is presented to and received from the user. Examples of UIAs are agents that maintain user profiles or customize the method through which search results are presented. Mediator agents work with both UIAs and Collection Interface Agents (CIAs). The mediators are in charge of tasks such as monitoring the progress of a query or deciding which collection a query from a UIA should access. Lastly, the CIAs define how the collections of documents themselves are accessed.

Such an architecture should allow for a digital library to grow as technology does, with new features implemented more easily through use of pre-existing agents. Complex tasks would be achieved through using a team of agents, each working on a specific part of whole. While the framework presented by Atkins is seemingly complete and well-defined, much of the implementation has not been fully thought through. Exactly how the agents would interface with collections or even other agents is not well defined. However, it does present a different approach to handling the future expansion of digital libraries.

## **2.2.2 Collaboration Over the Internet**

Agents are only one possible avenue to enhance the services provided by a digital library. Typically agents are seen working only in collaboration with their one user, rather than working to assist a collaborative effort of users. This leads us to examine what has been done with Internet-based collaboration, and how graduate students might be able to work together more efficiently through using NDLTD.

### **2.2.2.1 CoWeb**

One approach to collaboration on the Internet is for users to build a Web site focusing on one topic. Through the Web site, users would be able to read thoughts, opinions, find links to related materials, and contribute their own ideas [Guzdial, 1999]. Guzdial has been experimenting with this idea. He has developed a system known as CoWeb that allows students to create a Web site collaboratively over the Web. This is detailed in his paper “Collaborative Websites to Support an Open Authoring Community on the Web”.

The premise behind the CoWeb is that anyone can edit any page or create a new one through a simple mechanism, without any security or synchronization checks [Guzdial, 1999]. Such a system is extremely vulnerable to an ill-intentioned user, but affords great flexibility in exchange. When a user goes to edit a page, they use simple text formatting commands in the body of their text. Optionally, users that have knowledge of HTML can take advantage of it to create more complex pages.

CoWeb is basically a simple text database at heart. It stores the text for the pages in a database that anyone can edit through the Web interface [Guzdial, 1999]. All previous versions of a page are kept on file, so in the case of a vandal or some simple error, a page can be restored to a previous state. The CoWeb software itself is written in the

Squeak programming language, and uses the Pluggable Web Server as its Internet host. The system itself is available on many different platforms, including Macintosh, Windows, and SunOS.

Many different uses for the CoWeb have emerged. The first use was for faculty to maintain a simple Web site for their course [Guzdial, 1999]. This was especially useful for those faculty who weren't as comfortable with editing and uploading files. It was used as a student assignment hand-in site for classes where being able to see the work of others could lead to a discussion or new ideas. While this may not be appropriate for all classes, since the student solutions would be viewable by all, some classes used it for posting "model" assignments after grading, as a reference for the class. Other uses were for collaborative writing projects, design reviews, and focused discussions.

However, several uses came about that were unintended. One example of this, as Guzdial explained, was when a student started a "Choose Your Own Adventure" like book about one of the assignments for the class. This was completely unsolicited by the teacher, and was created by a student as a fun distraction on the CoWeb. Practically three dozen pages were created for it by a group of students in the class.

The best use for the CoWeb is to assist learning through the creation of a public artifact. It is not designed for non-educational activities as it is simply too inefficient. The CoWeb encourages users to get involved through reading whole pages of text and then making comments. While it was shown that it can support collaborative writing, it is not the optimal environment, for large documents. However, the CoWeb is easy to use and seems to work well in educational settings. This paper however did not make any attempt to assess what actual benefits are gained through the use of CoWeb. There was no usability study conducted, nor was there an in depth analysis of the usage of the CoWeb.

#### 2.2.2.2 Effectiveness of CoWeb

In order to gain some idea as to the effectiveness of the CoWeb, a survey was conducted to assess students' reactions [Guzdial, 1999]. The study attempted to examine the actual benefits of using a CoWeb, how students perceived it, and how it compared to other similar technologies. The surveys were distributed to volunteers in three classes that utilized a CoWeb: a Biology class, Chemical Engineering class, and Computer Science class.

The survey consisted of approximately 20 questions with Likert scales [Guzdial, 1999]. One limitation is that the survey was a preliminary one and that a better one could be designed for future research. There are no questions on the survey to verify or triangulate a concept, which reduced the reliability of the results.

The results of the survey showed that students found CoWeb quite easy to use [Guzdial, 1999]. No one stated that they had problems reading or editing pages with the system. “Students also found it useful, they liked using it, they want to use it in other classes, they weren’t frustrated by it, and they didn’t find reading and writing a chore” [Guzdial, 1999]. It is also interesting to note that students perceived that the CoWeb helped them in learning and completing tasks for class.

Lastly, the survey provided some of the first insight into the perceptions of security and trustworthiness of CoWebs. Students who created pages were actually more trusting of information in the CoWeb, but they were also the most bothered by the fact that anyone could edit any page [Guzdial, 1999]. It was conjectured that this might be because the time they invested in creating a page could be undone by anyone.

Something similar to the CoWeb could be of use with a collection of ETDs. Graduate students could browse not only the ETD itself, but an entire Web site devoted to it. The Web site would be an open forum for people to contribute. The negative aspect of this approach is that it lacks structure, and would most likely be intimidating to users when little or no information is contained in the Web site. A variation upon this idea, with more structure already provided, might be better.

#### **2.2.2.3 CoNote**

One tool has been developed to utilize annotations as a means of collaborative learning. This tool, named CoNote, is designed to share annotations within a group for a set of documents. It was developed for use with college classes, in place of a newsgroup-based discussion. It allows document authors, or the person who makes a document available, to select points throughout the text where annotations can be inserted. Students can come along and start discussions based on the context of the document at that particular point. While CoNote is a collaborative learning tool, it also bears a lot of similarities to the one that was developed for this study. Therefore, it is discussed further in section 2.3.3.1.

### **2.2.3 Summary of Areas Related to Possible Digital Library Tools**

There are many different possible tools that could be developed for use with digital libraries. Agents can provide a great number of powerful operations. However, they require a set of knowledge or an ability to learn, both of which are costly to program. On the other hand, the Internet provides an excellent medium for collaboration and

discussion. While this requires little artificial intelligence programming, there are issues of synchronizing data and providing the right amount of structure. This study looks to expand upon the body of knowledge on collaborative learning via the Internet. It will present a tool for groups to work together in finding information relevant to their research. This will be accomplished through annotations.

## 2.3 Annotations

In order to properly design an annotation system for NDLTD, further exploration was needed of prior related work. The subject of digital annotations has been well studied. Papers reviewed fell into two groups, application-based and Internet-based annotations. Application-based annotations are programs, such as Microsoft Word, that support annotations on local documents. Typically these systems are designed to annotate only a single document at a time. An Internet-based system is based upon using the World-Wide Web and browsers to annotate a large collection of documents and easily share those annotations. Table 2-2 shows the characteristics of many common methods of annotation, and how they differ from each other. Before delving into digital annotations, traditional annotation methods will be briefly mentioned.

**Table 2-2 Comparison of Annotation Systems**

System	Easily Shared with a Group	Easily Shared with Public	In-Place Annotations	No Special Equipment	Searchable Annotations	Separable from Document	Supports Links
Adobe Exchange	•	•	•	•			•
ComMentor	•	•	•		•		•
CoNote	•	•	•		•		
Conventional Paper Annotations			•	•			
ETD Annotation System	•	•		•	•	•	•
Lotus Notes	•		•	•		•	
Microsoft Word	•		•	•			•
XLibris			•		•		•

### **2.3.1 Paper-Based Annotations**

The notion of annotating a document has probably existed since the invention of writing. It is a natural behavior for most people, and an idea that is commonly understood. Today there are many methods through which a document is annotated. On paper, annotations can be written notes in the margin, highlighted text, corrections, or a summary. Writing annotations on a paper document is quick and easy. All you need is the document and a writing implement. However, there are drawbacks to the traditional method. In order to share your annotations, you must pass the document to another person. To share with a group, you must either make copies or designate a “passing order”. With multiple copies of the document, combining the annotations of others onto one copy can be a time consuming task. Another limitation is that you must physically have your annotated copy in order to refer to it. This is not necessarily a limitation of Internet-based digital annotations which you can access from an Internet-connected computer. The concept of annotating a document electronically aims to solve the problems with conventional paper-based annotations.

### **2.3.2 Application-Based Annotations**

Most document handling applications on the market today, such as Microsoft Word and Lotus Notes, have built-in facilities for annotating a document. Lotus Notes provides a scheme for annotation through providing “hooks” where annotations can be attached. Microsoft Word supports annotations through its “Track Changes” feature. It is primarily designed to make copy editing in Word easier by showing what changes a person suggests. However, through this facility, it also provides the ability to add comments. Adobe Exchange allows Post-It like notes to be placed onto a document. However, the free Adobe Acrobat Reader does not support writing annotations on documents, only viewing notes already in the document.

All three of these systems store the annotations with the document itself in their own proprietary document format. This makes sharing annotations easier, as you can just send the document to another person via email and they’ll have all of your notes. Once annotations are added to the document, you are forced to distribute the document with them until you delete them; something which may not always be desirable. It is difficult to combine the annotations of others, as little or no support exists for merging two versions of the same document. It can be difficult to have annotations which are private, and others that you want to share, all in the same document. One drawback is

that you must have a computer in front of you in order to read or write annotations. Xerox has taken a different approach, by creating an electronic reading device that looks to improve upon the approach of Microsoft and Adobe.

#### 2.3.2.1 Review of Introducing a Digital Reading Appliance

Xerox PARC (Palo Alto Research Center) has created a digital library reading appliance called XLibris [Marshall et al., 1999]. As Marshall explains, this reading appliance is designed to reproduce the feel of working with a paper document. This is accomplished by creating what is essentially a cross between a pen-based notebook computer without a keyboard and a legal pad without paper. Users can do all the same actions that they can with a paper document, such as highlighting, drawing, and writing. Other actions are also possible, for example XLibris can extract clippings based on annotations and place them into a Reader's Notebook for review.

A study was conducted by Marshall with a group of seven researchers who regularly read papers and annotated them. The goal was to determine to what extent a digital annotation device would be utilized and how people would react to it. However, other interesting results emerged about how people read and annotate. The study revealed that the researchers typically skimmed and re-read, referenced a previous page for a few seconds, as well as deliberately taking breaks even though the papers were short [Marshall et al., 1999]. Most of the researchers thought that XLibris was surprisingly "paper-like".

Even with the paper-like feel of the appliance, readers found that the margin sizes were too small for marginalia; this leads to the conclusion that larger margins promote annotations [Marshall et al., 1999]. More interestingly, the researchers did not get much use out of the Reader's Notebook feature, which summarizes their annotations, because during group discussion it was often more beneficial to see their annotations in the full context of the document. However, most of the group thought that the idea of keeping clippings over a period of time would be useful to come back to for reference.

In conclusion, the study revealed that their readers are mobile and read when the opportunity presents itself. The readers also desired the ability to extract key points from a document, to be placed into another document for future reference. XLibris does have its limitations, however. One of the largest drawbacks is that it is not designed for collaborative work [Marshall et al., 1999]. Another consideration is that you have to own an XLibris appliance, which, even after it is out of the prototype stage, may be too costly for the average graduate student. While XLibris is not the method of choice for annotations with NDLTd, it does provide some good indications as to what features

users expect from a digital annotation system. In the next section, we examine some Internet-based annotation systems to see what has already been developed for the World-Wide Web.

### **2.3.3 Internet-Based Annotations**

With all of the information contained in the World-Wide Web, the need to recall what you saw and where you saw it is of great importance. One method of accomplishing this is through annotations, a feature that is practically non-existent in the latest Web browsers. Researchers, realizing the need for such a capability, have set out to determine the best method for implementing annotations for the Web. Most of the systems developed thus far require an enhanced browser or special Web servers in order to combine pages and annotations. The CoNote system is one that requires a special server.

#### **2.3.3.1 CoNote**

CoNote's special Web proxy server works by processing pages before presenting them to the user [CoNote]. The person who makes a page available on the server can identify points within the document at which he/she believes annotation would be useful. When a user requests the page, the Web server will insert links at those points that will lead to a page that lets users write annotations in a threaded discussion format.

There are several drawbacks to CoNote's design. First, no privacy or security is implemented. All the annotations that are written are publicly viewable [CoNote]. Another major limitation is that annotations can only be written at the locations pre-specified by the person making the document available. This increases the cost of annotating a document, as someone has to spend time identifying the points at which annotations might occur. It would be difficult to anticipate all of the possible points.

Through a careful study of several Computer Science classes, CoNote was shown to increase participation in online discussions [CoNote]. This is a good indicator that Internet-based annotation systems are a beneficial tool. The one greatest benefit to the end-user, though, is that they do not need any special software, which increases the likelihood that they'll take advantage of it. Little work is required to utilize the system. This is unlike the ComMentor system, which requires users to browse the Web with a tailored version of the NCSA Mosaic browser [Roscheisen et al., 1995].

### 2.3.3.2 ComMentor

Through the use of a specialized browser, ComMentor eliminated the need to pre-identify places for annotation [Roscheisen et al., 1995]. ComMentor also worked to address some of the other limitations of a system like CoNote. One improvement was the concept of “annotation sets”. Every annotation belongs to a set. The user is able to tell the browser which set of annotations they want to view with a particular page. This provides for a layering-like effect, where the user could turn on their private annotations as well as the annotations of a research group. Annotation sets can be made publicly available, allowing anyone interested in a particular set to view what was written, and possibly contribute to it.

With ComMentor, annotations within a document are denoted by a small icon at the place where the annotation was written [Roscheisen et al., 1995]. The icon identifies the author of the annotation, which enables a user to immediately identify who wrote what. While viewing a particular annotation, responses to it are also available, allowing for some discussion to take place. Links to other documents or annotations can be inserted in an annotation in order to help make logical connections that don’t already exist. Annotations can be searched on a number of fields, such as author and date, which simplifies the task of finding your annotations, or the new ones added since you last visited.

A drawback to ComMentor is having to use both a special browser and server. The browser has to be specialized in order to be able to communicate with the server and display only the desired annotation sets to the user. It also has to be specially designed in order to eliminate the need to pre-determine the locations for annotation within a document. Using ComMentor is not as attractive to users since they have to abandon their browser for one that supports this annotation system.

### 2.3.4 Summary of Annotation Literature

Annotations on the Web have been shown to be useful. Systems like CoNote and ComMentor, while very useful, have limitations that are not acceptable in an annotation system for ETDs. It would be a laborious task for an author to predetermine places for annotation within their thesis or dissertation. Downloading a special browser would discourage many users, especially those who might be less computer savvy. Therefore a system that doesn’t require special servers or browsers is needed for the NDLTD Annotation System. This can be provided by forfeiting

the ability to have “in-place” annotations, the ability to write an annotation in the document near the context to which the annotation pertains.

Before deciding to implement an annotation system, a preliminary study was conducted of Computer Science graduate students and faculty. The goal of this study was to investigate which tool might be the most beneficial to graduate students. The next chapter presents this preliminary study and its results.

### 3 Preliminary Study

#### 3.1 Introduction

Before any tools could be developed for the digital library of ETDs, a preliminary study was needed to assess the usefulness and feasibility of a variety of possible tools. A brainstorming session was scheduled with the Virginia Tech ETD project team. The group came up with a list of nine tools that might be of benefit to graduate students, and which provided capabilities not easily achievable with a traditional paper thesis or dissertation. The ideas generated from the brainstorming session are summarized in Table 3-1.

**Table 3-1 Possible Digital Library Tools**

Bibliography builder	Compare & Contrast two ETDs	Jargon Definer
Find open problems	Influential article search	Find common approaches
Define a syllabus by concepts	Different ways to present a concept	Annotations

The list of tools was turned into a set of scenarios, each describing a person trying to accomplish a particular task. However, the scenarios were carefully written to leave out any detail as to how that task would be accomplished, other than that they were to involve a digital library tool. The goal was to have participants in a focus group visualize how these tools, as well as others they might think of, would work and what results they would produce. With the set of scenarios and a group of volunteers selected, the focus group was ready to take place.

#### 3.2 Procedure

The focus group was held in a lab specially designed for collaborative work via computers. Each workstation was setup to be on a Web page listing the scenarios. A copy of WordPad was launched so participants could write their visualizations of how the digital library tools presented in the scenarios would work. Instructions were given verbally as well as on a large projection TV. The participants were solicited from Computer Science graduate students and faculty members because it was thought that they would have the least trouble visualizing how these new tools might function in a realistic manner.

The group was asked to read a scenario and explain how the tool described would function. (The scenarios are shown in Appendix C). Participants were told that they could create their own scenarios and visualizations as well. Careful instructions were given to stay away from describing the user interface of the tool, but to instead focus

on the overall concept and how the results would be formatted. One of the goals of the focus group was to determine what the tools needed to contain in order to be usable. Thus it was necessary to gain an understanding of what type of input and output users expected would take place. Another benefit of the focus group was that it provided “fleshed-out” concepts of the original ideas from the brainstorming session in a quick and efficient manner.

As each participant completed a scenario, their ideas were posted onto the Web site for others to view. All participants were expected to periodically check the work of others to make comments and give a different perspective. Unique ideas would be brought up on the projection TV for the whole group to see and discuss out loud. This was done to encourage a collaborative effort and to ensure that the complex issues behind each tool were thoroughly thought out.

### 3.3 Results

The focus group produced some interesting and unexpected results. The visualizations of most tools were fairly similar, with a few exceptions. Generally, there would always be one participant with a unique perspective that would generate a good discussion among the group about the particular tool. Though the study did not make use of any software package designed for collaboration in focus groups, the amount of interaction was good.

The scenarios for the tools that compared two ETDs and that would find influential articles were two that most participants seemed most interested in. However, there was no tool that stood out as being the most useful. The visualizations of each tool were in-depth and provided insight into how people would like to take advantage of the added capabilities of a digital library. The ability to easily work with and cross-reference multiple documents at the same time was perceived as a desirable feature. Another trend was noticed in the area of recalling what had been read, the ideas it had generated, and what possible value it might have. This was particularly evident in the visualizations of the annotation scenario, and also surfaced in one participant’s idea of an “idea tracker”.

### 3.4 Conclusions

It can be concluded that the participants saw real value in the possibilities of digital library tools. They saw how some of these tools, which were unthinkable with paper theses, could save them time throughout all stages of their graduate education. This was encouraging in that it provided some evidence that further research into digital library tools would be beneficial. It was disappointing that the preliminary study was not able to identify one or two tools as being the one that students would like to have the most. However, this problem could be solved by

presenting the results to the NDLTD Steering Committee during their Fall 1998 meeting, to receive recommendations on where to focus future research.

Their recommendation was to develop the annotation system first, as they thought it would not only provide graduate students with a method to track their research, but also improve scholarly exchange between all individuals who use NDLTD. An annotation system would provide the ability to carry on discussion in the context of a particular thesis or dissertation, as well as a way to record personal thoughts. This system was implemented and ultimately became the basis for the usability study discussed in chapter 5. The next chapter describes the design of the annotation system itself.

## **4 The Annotation System**

### **4.1 Introduction**

After the decision was made to implement an annotation system for NDLTD, a plan was developed. This plan consisted of the following stages:

1. Develop a set of usage scenarios and define groups of potential users.
2. Determine how annotations will be stored and connected with their respective documents.
3. Define how access to annotations will be handled.
4. Implement the system using Perl scripts.
5. Test and evaluate.

In completing the first stage, several groups of users were identified. The primary group would be graduate students, while other groups included faculty, industry researchers, and thesis committee members. Scenarios were written that defined the different actions that could be performed with the annotation system, as well as outlined the motivations for a user to use the system. Some of the possible uses for the system were to: keep track of personal notes about ETDs, allow the author to post new contact information or publication information, and allow research group members to post links to relevant works published by them.

From the scenarios, a basic list of actions was derived, namely the ability to write: public annotations, private annotations, and those that would be shared with a particular group of users. Furthermore, users could read annotations, search by particular fields, and have a list of all the annotations they had written. These actions had to be taken into consideration, along with a number of other factors, throughout the design process.

### **4.2 Design Considerations**

The goal was to produce a system that would maximize usefulness and minimize the work required by the user. In order to minimize what the user has to do in order to take advantage of the annotation system, it was immediately obvious that making them use a specialized browser or connect to a special proxy server would be asking too much. Therefore it was decided that an annotation system would have to be developed that utilized the capabilities of an existing piece of software, such as Adobe Acrobat Reader or the GGI functionality of a Web server.

An investigation was conducted into the annotation features of Adobe products, since most of the ETDs are viewed from the PDF format. It was quickly discovered that annotations could only be written with Adobe Exchange, which would cost the users money, and that the annotations become part of the document itself. This was a serious limitation, as it would necessitate keeping separate versions of the ETD, because each annotation set had to be contained inside of the ETD file. Not only is this inefficient in terms of storage space, but it makes layering several sets of annotations very difficult. A search was conducted for third-party plug-ins to Adobe Acrobat Reader that implemented annotations stored in a separate file, but none was found.

This left only two options, the construction of a plug-in for Acrobat Reader specifically designed for annotations on NDLTD, or building a set of routines to offer annotations through Web pages. Due to the fact that formats may change over time, it was thought that the Web routines method would create a longer lasting annotation system. While it is highly desirable to have annotations in-place in the document, this is not possible with this solution. Instead it was decided that the annotation system would be entered from the abstract page of a particular thesis or dissertation. This way annotations still have a connection to their document, but they can be ignored by those who do not wish to use them.

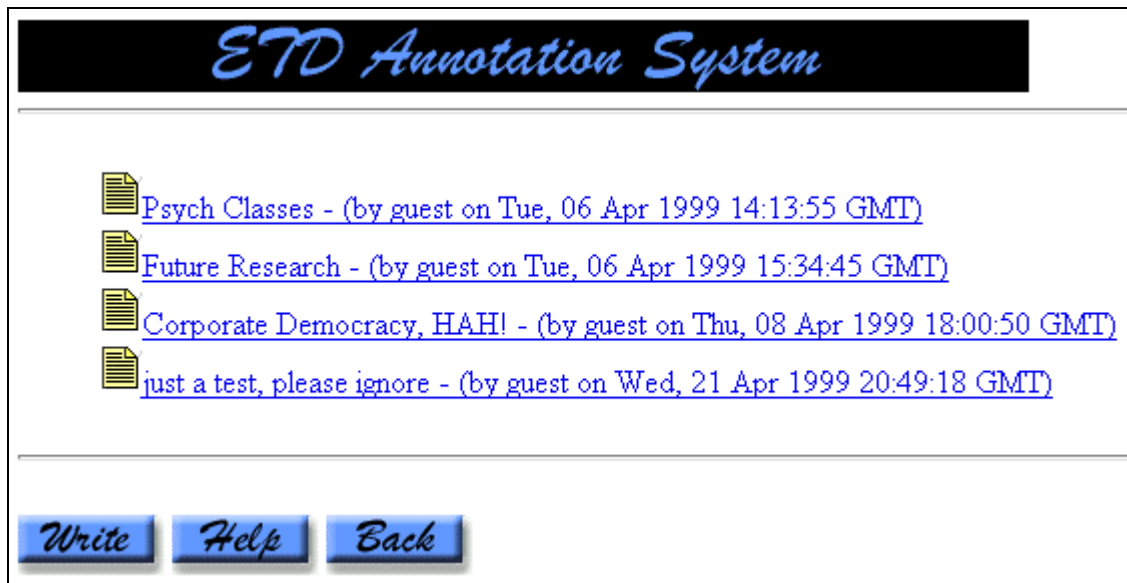
The next major issue that was tackled was the topic of security and access. Three different types of annotations had to be provided: public, group, and private. A great deal of discussion took place that focused on who could write public annotations. Some thought that it opened up the possibility for someone to post inaccurate or slanderous material. Others thought that allowing anyone to post publicly would promote a higher level of scholarly exchange. There was also the question of whether posts are anonymous or identified with their authors. In the end, it was decided that only those people authorized by the author of a ETD would be allowed to post publicly, and that all posts would be identified with their authors.

Now that most of the design issues had been tackled, the task of implementing the system remained. System design is covered in the next section, while details on how the system actually works are described in Appendix A.

## 4.3 Implementation Overview

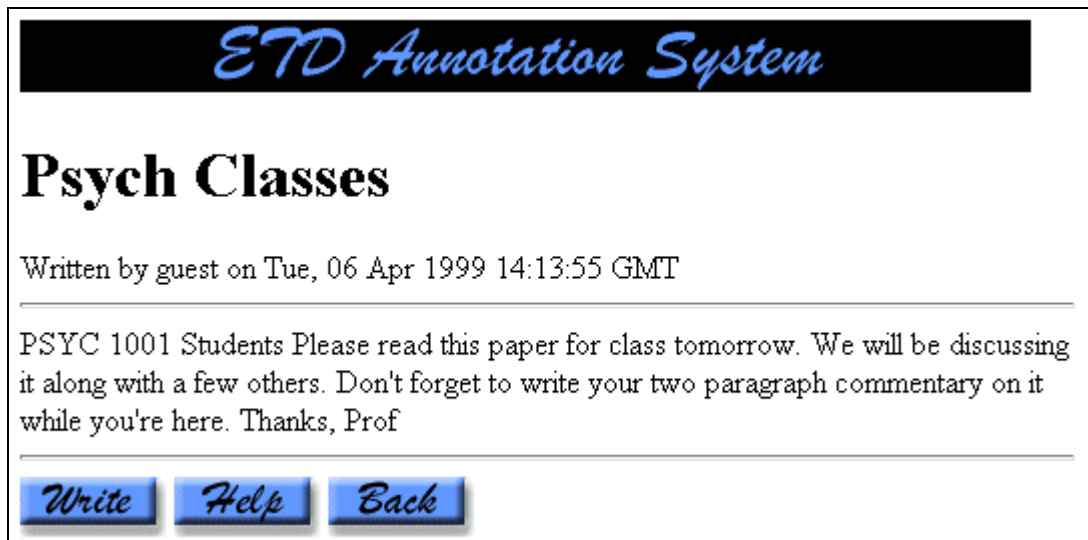
The annotation system was designed to be intuitive and easy to use. A user either browses or searches the collection of ETDs available. Upon finding one of interest, they select the ETD and receive its abstract page with links to the document itself, along with a button labeled “View/Write Annotations”. The user can click on this

button, at which point they'll be asked to login using their Virginia Tech PID and password if they haven't already logged in. Once they have been verified, a list of annotations that they have access to read is presented. This list, as seen in Figure 4-1, shows a subject, the author's PID, and the date on which the annotation was written. New annotations appear at the top of the list. There is no threaded structure to the list of annotations, even though responses to previous annotations may exist. The user is presented with the ability to read or write annotations, receive help, or go back to the abstract page.



**Figure 4-1 Screenshot of List of Viewable Annotations**

Should the user decide to read annotations, they can click on the subject of the annotation they wish to read and its text will be presented. One example of an annotation is shown in Figure 4-2. At this point, they can choose to write a response, go back to the list of annotations, or receive help. A response is treated as a new annotation, but with the subject already filled in with "RE:" added as a prefix.



**Figure 4-2 Screenshot of Reading an Annotation**

When writing a response or new annotation, the user can specify their subject and an unlimited amount of text. The form used to create new annotations is illustrated in Figure 4-3. The user then has up to three options when specifying the access permissions of their annotation. If the user has been identified as someone allowed to post publicly, the option to make their annotation public to all is provided. Otherwise, all users can either specify a list of PIDs with which they'd like their annotation to be shared or they can mark it as private, which enables only them to see it. The PIDs of users to share an annotation with are the same PID as their email address, which can be easily found through the Virginia Tech People Finder.

# ETD Annotation System

## Add an annotation

Write your annotation below, then select a publication option.

Title:

This thesis contained some interesting ideas pertaining to the development of new studies.

It could be beneficial to use the instrument developed for this particular experiment. However, the same limitations may arise from it.

Publication Options

☒ **Private** (only you can view it)

☐ **Selected users** (only PIDs entered in space below can view it)  
Separate PIDs with a space

☐ **Public** (everyone can view it)

Add Annotation

Figure 4-3 Screenshot of Writing an Annotation

Context-sensitive help is available on all screens to give the user guidance should they become confused. Despite the fact that a user must log-in to use the system, there is no method to log-off; users simply go to other pages on the World-Wide Web. After approximately five minutes, their log-in will expire and further access to the annotation system requires them to log-in again.

## 4.4 Future Work

Two features are currently in development. One of them is the ability for users to go directly back to the abstract page at anytime. The second feature is a special Web page that is available to users, that would list all the annotations they've written. It will serve as a quick reference to their notes, saving them the effort of working through the ETDs they've read, to find their annotations. There are other areas that could be further developed with

the current system. One feature that would be of great benefit is the ability to search for new annotations that have been added since a user was last on. Another is the ability to perform a keyword text search of all the annotations viewable by a user. Also valuable would be a method for groups receiving annotations to be more easily defined, to avoid having to enter in a list of PIDs to give access to a group. This would make the system easier to use with groups of students, and also provides a starting point for creating layers of annotations that can be turned on and off for viewing. Lastly, allowing people to apply for public annotation abilities might be beneficial in increasing the utility of the annotation system and the scholarly exchange that can take place through it. The next chapter will cover the usability study, which looks at what students thought of the current annotation system.

## **5 Summative Evaluation**

The purpose of this study was to examine the usability and usefulness of the annotation system developed for NDLTD. How students perceive the system, what they see as its potential uses, and what would encourage them to use it are all key factors. The system needs to be as easy to understand as possible, yet providing all of the features that graduate students would desire. This study set out to verify that a beneficial tool had been developed and that students would actually use it should it become available for general use. It also looked to identify any oversights in the basic design that might make the system difficult or frustrating to work with.

Specifically, the study was designed to answer the following three research questions:

- 1) Do graduate students find the annotation system useful and easy to understand?
- 2) How would graduate students anticipate using the annotation system?
- 3) What features would make the system more beneficial to graduate students?

### **5.1 Sample Selection**

Due to the focus of improving graduate education, only graduate students were asked to participate through an email sent out to all graduate departments. Perspective participants were informed that the study would be examining the usefulness of the ability to annotate the collection of ETDs. It was decided that the study would be open to all graduate students at Virginia Tech in order to avoid any bias based upon field of study, degree, or research progress. No rewards were offered to participants other than free food. Thus the sample was purely voluntary, based upon their interest in the new technology.

### **5.2 Instrumentation**

Since the annotation system developed was the first of its kind, there were no preexisting instruments that could be used to study its usability. The decision was made that a questionnaire divided into several parts would be the best method to record the experiences of the participants. Videotaping the volunteers was considered, but was not done because it did not seem likely to lead to a substantial improvement in results.

The questionnaire (shown in Appendix B) was divided into four parts, and assumed interactions with the system in between parts. The first part covered basic demographic information and tried to assess the participant's research and computer use habits. After completing the first part, the volunteer would interact with the annotation system by reading several of the annotations that were already present. Once a sufficient amount of time had

elapsed, the participant would be asked to complete the second section of the questionnaire. This section asked questions about their impressions, such as what problems they experienced and their overall reaction. Before proceeding to the next section, a participant was asked to interact with the system, this time writing two annotations. After writing their annotations, the volunteer would again answer questions about their experience, gauging how easily they were able to accomplish this task. Once the third section was complete, the participant was told that they could explore the system at their leisure and then answer the last section. The last section asked questions geared towards determining how the participant saw the system being used, what their use of it would be, and what would make it more attractive to use.

Since this instrument had to be developed from scratch, it was decided that a pilot study should take place to determine the best procedure, as well as to discover any oversights in the original questionnaire. The pilot study was conducted with a group of volunteers, solicited in the same manner as the participants for the actual study. Three graduate students participated. Based upon their comments and results, no major changes were called for before conducting the actual study.

### 5.3 Data Collection Procedures

Groups of three volunteers were formed based upon their availability. The groups were told to meet at the usability labs in one of the buildings on-campus at the specified time. The study was conducted in the same collaborative lab as the preliminary study, with each participant working at their own workstation and instructions being displayed on the large projection TV. As group members arrived, they were given a consent form and asked to sign it. Once consenting to participate, the questionnaire was given to them.

Participants were allowed to complete the first section of the questionnaire while the group arrived. However, no one was allowed to proceed any further until everyone was ready. Instructions were given verbally, re-explaining the purpose of the study and the procedures for it. Participants were then asked to use the annotation system to read annotations for five minutes. After the time expired, they were instructed to answer section two of the questionnaire. Once everyone was ready, instructions for the next section were verbally given and participants were allowed ten minutes to write two annotations and then complete the third section of the questionnaire. Once participants completed the questionnaire, they were instructed to freely explore as they felt necessary and then complete the last section of questions.

## 5.4 Limitations

The nature of sample selection is a possible limitation of this study. Due to the participants being completely voluntary in nature, their perceptions may vary from those who chose not to volunteer. This may be compounded by the fact that the usability study only involved graduate students. The sample sizes for both the focus group and the usability study were small and could affect the results, in that a larger sample may be able to provide a more accurate “average”.

The study has been limited by the characteristics of the participants. In the usability study, only graduate students participated. Further, while students at different points in their studies volunteered, the data was only collected from them once, as the study made no attempt to track changes in their perceptions over time.

The study has been restricted by the instruments used to assess the perceptions and ideas of the students and faculty. The scenarios provided in the preliminary study may not have been sufficient or properly posed to have successfully generated the proper feedback; these scenarios were not derived from any previously proven source. It is possible that the usability study’s questionnaire may not have produced accurate results, as another instrument might have provided conflicting data. The questionnaire also was not derived from any previously proven source, and had to be designed from scratch.

Finally, this study also may have been limited by the fact that only graduate students and faculty at Virginia Tech participated. It is possible that their views could differ from those of other students and faculty at other universities.

## **6 Results**

In order to examine the benefits gained by digital library tools, participants in the usability study completed questionnaires that recorded their experiences with the annotation system. When only two student volunteered for the actual study, their results were grouped with the pilot study to form a sample size of five. In order to ensure that more volunteers were not necessary, the amount of variance between answers was compared and subjectively found to be small enough to not warrant another group of participants.

Results of the data collected are provided in this chapter, divided into two parts. First, the five participants in the study will be described. The second section will present a summary of the results from the questionnaires.

### **6.1 Description of Participants**

The demographic information asked in the questionnaire revealed a diverse sample of graduate students. 40% of the participants were female, and 60% were male. While mostly masters level students participated, one doctoral student was present. All participants had begun their research, some further along than others. There was a wide range of self-assessed computer expertise, from moderate knowledge (5 out of 10) to expert (9 out of 10). The average expertise among the group was a 6.8. Three participants used the Web on a daily basis, one reported only weekly usage, and another reported using the Web several times a day.

As far as research habits, roughly half of the participants reported regularly writing summaries or annotations when reading, whereas the others reported not doing so, or only on occasion. Four out of five participants reported having trouble remembering where they had previously read things and the ideas that were generated. Most participants agreed that managing large quantities of data and research materials was one of the biggest difficulties in conducting research.

### **6.2 Summary of Responses**

Due to the low number of volunteers for the actual usability study, the data was compared to that of the pilot study. The variance of answers was small enough that it was seen fit to combine the two groups of data. This created a larger sample size of five students, which increases the reliability and validity of this study.

The responses for participants after they had completed interacting with the annotation system to read pre-existing annotations were fairly consistent. No one reported problems using the system in this first stage and

everyone found it convenient to read an annotation. Most reported that they would be likely to look for annotations on documents when reviewing them in order to gain any additional perspectives or information. One of the interesting comments was, “Very convenient – I appreciated the date/time stamp so I knew how recent it was.”.

In the third section, again no one reported having any problems using the system to write annotations. Most agreed that they would be likely to utilize the feature to write personal annotations in order to remember what they found useful or intriguing about the document. One participant commented that they would only write annotations about half of the time and that they’d be more likely to write annotations only for private use. Several suggestions were made for improvements to the system at this point, including a page to view all annotations that they had written, a method to view all annotations at the same time, and the ability to jump back to the abstract page from any point.

In the last section about the annotation system in general, three of the participants commented that they liked being able to control the privacy of their writings. All but one participant agreed that they would find the system to be beneficial to them in conducting their research, and that reading the annotations of others would be helpful as well. One participant’s comment summarized most, “It could give other views of the work”. Everyone said that they would be likely to tell others about the existence of the annotation system and most would be likely to share their annotations with friends.

No one stated that the system was complicated and all agreed that writing annotations was easy. The comment “Good layout – easy to follow” was similar to most responses. Some comments were made on navigation issues. One participant noted that they weren’t sure if they should use the “Back” button on the Web page or the browser’s back button. Another one commented that the ability to jump back to the abstract page was needed. On a scale of 1 to 10 with 10 being very satisfied, the overall satisfaction with the annotation system was 7.9.

## 7 Discussion and Implications

Based upon the results from the questionnaires of the participants, some conclusions can be drawn. This chapter will present how these findings can be utilized by first showing how the study provides answers to the research questions. Implications of this study for future annotation systems and future digital library tools will be discussed in an attempt to provide some direction for future research.

### 7.1 Responding to Research Questions

The data collected provided valuable insight into answering the research questions posed by this study, as will be shown in this section. Even with the small sample size, the results were consistent enough to draw strong conclusions from.

The first research question asked if graduate students find the annotation system useful and easy to understand. From the questionnaires, it can be said that students found it both useful and easy to understand. This is clearly seen in the fact that all participants agreed that they had very little trouble using the system. The data also shows that 80% of the participants found the annotation system to be useful and that they would utilize such a tool.

This leads into the second research question of how would students anticipate using the system. Most respondents had the same perspective that they would use the system for two purposes. The first was to keep track of their own notes and thoughts after having read a particular thesis or dissertation. This would be done through private annotations they would write using the system. Second, they said they would utilize it through reading the public annotations to find out information on related studies or different perspectives on the document.

Lastly, the question was asked what would make the system more beneficial to graduate students. Everyone seemed to have some ideas on what would make it more useful, but two ideas were prominent. The first idea was to have a page to show all annotations that they had personally written. This way they could jump straight to where they had been, rather than having to relocate the ETD in order to find their notes. The other idea focused on the ability to search for annotations, whether it be by keyword or by date to find new ones written since they were last there. This feature would enable them to also find other annotations on documents that might be useful, but that they had not read yet.

## 7.2 Implications of the Study

This study has successfully answered all three of the research questions posed. It shows that an annotation system is useful and can provide abilities not possible with paper theses or dissertations. These results lead to some interesting implications for future research, in both annotations systems and digital library tools.

### 7.2.1 Implications for Future Annotation Systems

The study has revealed several factors to consider in designing annotation systems in the future. The first factor is that in-place annotations are not a necessity for people to find them useful. Most current annotation systems go to great lengths to provide in-place functionality, when keeping annotations in a separate area that is linked to the document is not only easier to implement, but also makes sharing and viewing sets of annotations a much simpler task.

Another implication for future Internet-based systems is that the need for a special browser may not be a successful solution. In this study the graduate students felt that they were only moderately knowledgeable about annotation, and most only used the Web once a day. Such users may not feel confident in installing a new browser unless they can readily see the benefits that it will bring and that the installation will be simple. Using Perl CGI scripts proved to be very successful for this annotation system, and it required no special Web browser or server capabilities.

The annotation system developed for NDLTD has been established as a successful model. Since it has been shown to be easy to use and beneficial, it could provide a good starting point for future annotation systems. The system itself is easily adaptable and can be run on any Web server that supports Perl 5. It may provide a long-term look into the usefulness of annotations on ETDs after the tool has received extended use.

### 7.2.2 Implications for Future Digital Library Tools

This study has shed more light on the development of digital library tools. It has been shown that graduate students find tools to be beneficial. Once they are informed of the added benefits of using a collection like NDLTD, they will most likely begin to use it more frequently to improve their efficiency in conducting research. Through using digital libraries and tools, graduate students will be more literate in electronic publishing and more likely to use it as a means of publishing their own work.

This study provides insight into what factors should be considered in designed digital library tools for graduate students. In this study, it became obvious that graduate students are most interested in tools that will help them better manage the large volumes of papers, articles, and theses that they read. The annotation system did this through keeping their notes online and by giving them the ability to quickly go back and see what their previous thoughts were. Therefore future research should consider how tools will enable graduate students to better manage the materials that they have to deal with.

It has been shown that NDLTD is beneficial to graduate students. Graduate students who had not used the Virginia Tech ETD collection, commented that they would have used it if they had known how useful it could have been to them. The benefits of having all the information on-line along with the tools that make using digital documents more convenient was readily realized by the students who participated in the study. While this study does not provide conclusive evidence that NDLTD raises the quality of graduate education, it does provide a starting point for future investigations.

## 8 References

- Lieberman, H. (1997, March). Autonomous Interface Agents. CHI 1997 Conference Proceedings on Human Factors in Computing Systems (pp. 67-74). Atlanta: ACM.
- Huttenlocher, D. (1994). Tools for Collaborative Learning. [On-line]. Available: <http://www3.cs.cornell.edu/dph/docs/annotation/annotations.html>.
- Kipp, N. A.. Flying Saucers, Wizards, and Deep Space: How we have applied HyTime in the Virginia Tech Digital Library Research Laboratory. [On-line]. Available: <http://csgrad.cs.vt.edu/~nkipp/meta98/paper/paper.htm>.
- Miller, J. S. (1996, November). W3C and digital libraries. D-Lib Magazine. [On-line]. Available: <http://www.dlib.org/dlib/november96/11miller.html>.
- Schatz, B., & Chen. H. (1996, May). Building large-scale digital libraries. Computer. [On-line]. Available: <http://computer.org/computer/dli/>.
- Atkins, D. E., Birmingham, W. P., Durfee, E. H., Glover, E. J., Mullen, T., Rundensteiner, E. A., Soloway, E., Vidal, J. M., Wallace, R., & Wellman, M. P. (1996, May). Toward inquiry-based education through interacting software agents. Computer. [On-line]. Available: <http://computer.org/computer/dli/r50069/r50069.htm>.
- Guzdial, M. (1999). Collaborative websites to support an authoring community on the Web. Journal of the Learning Sciences, Submitted. [On-line]. Available: <http://guzdial.cc.gatech.edu/papers/CoWeb/>.
- Guzdial, M. (1998). Teacher and student authoring on the web for shifting agency. Presented at AERA 99 Session: How can CSCL (Computer Supported Collaborative Learning) change classroom culture and patterns of interaction among participants? Available: <http://guzdial.cc.gatech.edu/papers/aera99/>.
- Fox, E. A., Eaton, J. L., McMillan, G., Kipp, N. A., Weiss, L., Arce, E., & Guyer, S. (1996, September). National digital library of theses and dissertations: A scalable and sustainable approach to unlock university resources. D-Lib Magazine. [On-line]. Available: <http://www.dlib.org/dlib/september96/theses/09fox.html>.
- Roscheisen, M., Winograd, T., & Paepcke, A. (1995, August). Content ratings and other third-party value-added information defining an enabling platform. D-Lib Magazine. [On-line]. Available: <http://www.dlib.org/dlib/august95/stanford/08roscheisen.html>.
- Roscheisen, M., Winograd, T., & Paepcke, A. (1995). Interaction design for shared world-wide web annotations. Denver: CHI 1995. [On-line]. Available: <http://Walrus.Stanford.EDU/diglib/pub/reports/brio-chi95.html>.

## 9 Appendix A – Annotation System Implementation

This appendix covers the details of how the ETD Annotation System was actually implemented. The system itself was written in Perl 5, with the exception of a utility written in C to validate users' PIDs and passwords. Perl was chosen for: its utility to quickly develop CGI programs and wide acceptance. All of the graphics were drawn using Adobe Photoshop 5.0. Hypertext pages were created with Allaire HomeSite 4.0.

The system itself is really quite simple. One large Perl script takes care of the entire system. When a user logs in, their PID and password is verified using the Virginia Tech Mail Server. A unique session number is assigned to them, and transmitted back as both a cookie and hidden form fields. The session number allows the system to know who a user is without retransmitting their PID or password with every transaction. Every time a session is started, a session file is created to store variables such as the user's PID or what action they're performing. Old sessions have to be erased by a cron job to avoid a pile up of session files. This comes from the convenience afforded the user of not having to worry about logging out.

Once the user has logged on and the session file has been created, the user is free to read and write annotations. The list of annotations shown to the user is determined by searching a database of all annotations, comparing the document the annotation is from to the one the user was viewing, and the viewing rights of the annotation and user. In order for an annotation to appear on the list that the user can read, it must be from the same document as the user came from and either public, viewable by a group that contains the user, or private and written by the user.

Annotations themselves are stored in a separate directory. Each annotation receives a unique number, and the file is simply the annotation number with ".ann" as a suffix. The annotation file is a plain text file with the text of the annotation. Details about the annotation, such as author and date, are stored in the annotation database. This database is what makes the links from documents to their annotation files. It is stored in the Perl DBF format for quick access.

## 10 Appendix B – Usability Study Questionnaire

This appendix presents the questionnaire that was used for the usability study.

### Demographics

Gender: ☐ Male ☐ Female

Class: ☐ Undergraduate ☐ Graduate (M.S.) ☐ Graduate (Ph.D.)

Are you actively working on a thesis or dissertation: ☐ Yes ☐ No

### Computer Usage

Have you used digital library resources (e.g., pdf files, online journals): ☐ Yes ☐ No

Have you used the VT Electronic Theses & Dissertation collection: ☐ Yes ☐ No

Have you submitted an ETD before: ☐ Yes ☐ No

How often do you browse the Web: ☐ Several times a day ☐ Daily ☐ Weekly ☐ Rarely

What are your primary uses of the Web: \_\_\_\_\_

How would you rate your computer expertise (1=know very little, 10=expert): \_\_\_\_\_

Please explain: \_\_\_\_\_

For how many years have you been using a computer on a regular basis: \_\_\_\_\_

### Research Methods

What problems do you have when managing all the information for your research: \_\_\_\_\_

\_\_\_\_\_

When researching, do you write summaries or annotations: ☐ Yes ☐ No

If yes, how many and what about? If no, why not? \_\_\_\_\_

\_\_\_\_\_

What would you think of the ability to read the annotations and summaries of others who read the same document as you? (professors, students, people in industry, etc.,...): \_\_\_\_\_

\_\_\_\_\_

Do you have trouble remembering where you read something: ☐ Yes ☐ No

Explain: \_\_\_\_\_

\_\_\_\_\_

In what ways do you use the web to conduct research: \_\_\_\_\_

---

What has been some of the problems you have with using the web for research (other than network problems, speed, etc.,...): \_\_\_\_\_

---

If you use the web, are there any sites you visit often? \_\_\_\_\_

---

Are you able to find the information you need better online or at a library? Why? \_\_\_\_\_

---

### Instructions for Reader Role

*(Make sure the web browser is already running and on the sample ETD page)*

Now you will play the role of someone who is browsing the Virginia Tech collection of electronic theses and dissertations. Using your web browser, look at the thesis abstract. Try out the “View/Write Annotations” feature on the page by reading some of the annotations that are there. You may try experimenting with the other features of the system, but do not write any annotations.

### Reader Questions

*(Answer this section only after having completed the reader role)*

How convenient was it to read an annotation? Explain: \_\_\_\_\_

---

Would you be likely to read annotations or summaries on-line with such a system? \_\_\_\_\_

---

Did reading the annotations change your opinion of the thesis? \_\_\_\_\_

---

What changes could be made to make you more likely to use the system to aid in conducting your research:

### Instructions for Commenter Role

*(Make sure the web browser is already running and on the sample ETD page)*

Now you will play the role of someone who has found some useful information in a particular thesis for their research. Using the “View/Write Annotations” feature on the abstract page of the thesis, write writing a few comments or summaries about the document. You may try experimenting with any other part of the system.

### Commenter Questions

*(Answer this section only after having completed the commenter role)*

How convenient was it to write an annotation? Explain: \_\_\_\_\_

\_\_\_\_\_

Would you be likely to write annotations or summaries on-line with such a system? \_\_\_\_\_

\_\_\_\_\_

What changes could be made to make you more likely to use the system to aid in conducting your research:

\_\_\_\_\_

\_\_\_\_\_

### Annotation System

*(Answer this section only after having completed both roles)*

Did you have any problems using the annotation system? If so, what? \_\_\_\_\_

\_\_\_\_\_

What did you find to be the best feature about the annotation system? \_\_\_\_\_

\_\_\_\_\_

Was any part of the system confusing or hard to understand? \_\_\_\_\_

\_\_\_\_\_

Would you find such a system beneficial to conducting research? Why or why not? \_\_\_\_\_

\_\_\_\_\_

What might make the system easier to use? \_\_\_\_\_

\_\_\_\_\_

What would make the system more useful for you? \_\_\_\_\_

\_\_\_\_\_

Do you think you would share your annotations with friends?      ☐ Yes      ☐ No

Explain: \_\_\_\_\_

Do you think reading annotations by others would be helpful?      ☐ Yes      ☐ No

Explain: \_\_\_\_\_

Would you tell others about the system?      ☐ Yes      ☐ No

Explain: \_\_\_\_\_

Is the system complicated?      ☐ Yes      ☐ No

Explain: \_\_\_\_\_

Was writing an annotation easy?      ☐ Yes      ☐ No

Explain: \_\_\_\_\_

Were you ever unsure of what to do next with the system?      ☐ Yes      ☐ No

If yes, when were you unsure? \_\_\_\_\_

How would you rate your overall experience with the system (1=very poor, 10=excellent): \_\_\_\_\_

Any additional comments? (use back if needed) \_\_\_\_\_

\_\_\_\_\_

## 11 Appendix C – Scenarios for Preliminary Study

This appendix presents the scenarios that were used in the preliminary study. Graduate students and faculty members were asked to write visualizations of how these systems would interact with the user. Other scenarios were generated by the participants that are not included here.

**Annotating Documents.** While Joe is reading ETDs, he likes to jot down ideas that come to mind and short summaries of the documents. Joe also likes to read the public annotations of others to find different perspectives. How does Joe go about writing an annotation? What is the process to read one or to find those that Joe previously wrote?

**Building Bibliographies.** Tim is finishing up his dissertation, and now needs to put together a bibliography of works cited from the ETD library. He has to cite five dissertations, all from the Virginia Tech collection. What steps does Tim have to take to have the system generate a bibliography for him of those five papers? What if Tim is selecting from ones used in prior related papers?

**Comparing & Contrasting.** Joan is conducting research on mockingbird migration patterns, and she has found two dissertations that pertain and are of similar structure. In order to quicken her research, it would be beneficial to be able to view both papers at the same time in a manner that lends itself to comparisons. How does the system present the papers? What features enhance the capabilities beyond that of two windows? How can the differences be best shown? The similarities?

**Jargon Search.** John is an undergrad conducting research for an English paper on modern coffee shop design. However, she encounters a fair amount of architectural jargon that she is unfamiliar with. It would be of great help if the system could find other references to these words, and present them in some context that can help convey their meaning. How do you envision this system working? Can the system identify the jargon automatically? Can the system suggest keywords for a thesis based on unusual word or phrase occurrences?

**Find Open Problems.** The first step in starting a dissertation is finding the topic, and Moe is having trouble finding a good open problem to solve. There is an extensive collection of ETDs on his area of interest, and he would like to system to identify open problems based on what is stated in the texts. What steps would he take to have the system identify potential problems for his research? Will the steps change if some comprehensive category classification system exists?

**Find Influential Articles.** While reading over some theses, Katie notices that several of them make reference to the same papers or authors. Since there are over two dozen papers on her topic, it would take too long to correlate all the references. It would be nice if the system could determine which papers and/or authors are cited most, and present the context in which they are cited from each paper. What would best enable Katie to do this? How can ISI data help? How can the NDLTD collection be best combined with ISI data for this?

**Find Common Approaches.** Liz, a chemistry masters student, is putting together some experiments and would like to research approaches taken before. There is ample material available on this type of experiment before in the ETD library, so she turns to that source for information. How could the system easily present her with paragraphs talking about the approaches taken in a range of papers? Can they be summarized according to the most popular and most successful?

**Assemble Syllabi by Concepts.** Sam is planning a new course that combines multi-variable calculus and differential equations. He has a set of concepts that he would like to cover, but isn't sure as to what order to present them in and which methods to present due to time constraints. What would the system be like that could help find the best order and methods to present this new class in? If he wants to plan another math course, could the system also identify what concepts to cover?

**Find Different Ways to Present a Concept.** A student in Cal's Physics class is having a lot of difficulty understanding Newton's Laws. Cal has explained the principles several ways, but has run out of ideas. Thus he comes to the courseware library looking for alternate methods to explain this concept. How would this system work

to find different methods of presenting Newton's Laws? Can it use a log/history of similar experiences of other students to help? Can it use a pool of multimedia illustrations and simulations to assist?

## **12 Vita**

Todd Miller was born on December 25, 1975 in Greensboro, NC. He attended Greensboro Day School for K-12 education, and then went to Virginia Tech. There Todd double majored in Computer Science and Computer Engineering, receiving Honors Baccalaureate degrees. While attending Virginia Tech, Todd co-op'ed with AT&T and Northern Telecom over a period of three summers. After graduation, Todd will begin graduate school at Georgia Tech, with the goal of a Ph.D. in Computer Science.