Multimedia, Hypertext and Information Access

A Proposal to the National Science Foundation

in response to NSF Solicitation 96-36

Edward Fox, Principal Investigator
Professor, Computer Science
Associate Director for Research, Computing Center
Virginia Tech

with subcontracts to:

Brian C. Smith
Department of Computer Science
Cornell University

Richard Furuta, John Leggett
Department of Computer Science
Texas A&M University

Lawrence A. Rowe
Department of Electrical Engineering and Computer Science
University of California at Berkeley

and in-kind cost sharing collaboration from:

Scott M. Stevens
Department of Computer Science
Carnegie-Mellon University

October 8, 1996

PROJECT SUMMARY

Multimedia, Hypertext and Information Access (MHIA) covers some of the most technologically important areas in the broad field of information technology, computing and communications. However, there is a serious lack of trained personnel in this field,and relatively ineffective transfer of research results into education. This proposal involves investigators at Carnegie-Mellon, Cornell, Texas A&M, Univ. California at Berkeley, and Virginia Tech working with ACM, its Special Interest Groups in this area (SIGIR, SIGLINK, SIGMM), and IEEE Computer Society Technical Committee on Multimedia Computing and Systems. Through workshops related to these groups, curricular recommendations will be discussed, leading to guidelines similar to those developed in 1992 by ACM SIGCHI. These will be disseminated through SIG and society publications as well as presentations. At the same time, research in the MHIA field at these the 5 universities will be packaged to aid in education, and to fit in with the curriculum. A distributed digital library will serve as a courseware repository for the field, including also packaged versions of demonstrations, prototypes, and publications presented at the annual conferences of the societies and their groups. MHIA-related technology being developed at the 5 universities will be applied to local courses, collaborative distance taught courses, and hundreds of knowledge modules that will result from the courseware packaging work.

Carnegie-Mellon will extend its work with digital video and digital libraries. Its Entertainment Technology Center is developing an interdisciplinary masters. Cornell with work with its Rivl language, and extend its WWW catalog of course-related sites in the multimedia field. Texas A&M will continue development of its hypertext bibliography and digital library efforts, helping with hypertext-related curricular discussions. University of California at Berkeley will continue its work with digital video and its on-demand delivery for educational needs, relating that to digital library activities. Virginia Tech will extend its information retrieval, digital library and courseware efforts that began with an NSF Educational Infra-structure grant on “Interactive Learning with a Digital Library in Computer Science” and is extending to build a National Digital Library of Theses and Dissertations.

Corporate partners include Bellcore, Knowledge Systems, MERL, and others.

C. Project Description

Motivation, Overview: Multimedia, hypertext and information access (MHIA) covers some of the most technologically important areas in the broad field of information technology, computing and communications. In this field, enormous investment has been made in the last decade to develop faster networking, tailored computer systems, usable authoring software, integrated digital libraries, edutainment packages and a wide variety of applications. Low cost storage and ubiquitous networking have moved electronic publishing, hypertext access, search methods, and multimedia access into the main stream of the rapidly growing World-Wide Web. Tens of billions of dollars of our economy relate to these technologies that now underly the Information Age and flow over the Information Highway. Some of the most innovative research work in the science/engineering world deals with solving hard problems in compression, processor design, computer interfaces, and communication with acceptable quality of service. However, there is a serious lack of trained personnel in this field, and relatively ineffective transfer of research results into education. This proposal addresses the problem through a multi-institution team, led by PI Fox at Virginia Tech, that will combine research with education, and develop comprehensive curricula and a repository (digital library) of courseware - using a number of innovative concepts.

C.1. Results from Prior NSF Support

C.1.1. Details

Award Number: NSF CDA 9312611
Amount of Award: $449,088
Period of Award: 8/15/93 - 1/31/97
Project Director: Edward A. Fox
Co-PIs: JAN Lee, Clifford Shaffer, H. Rex Hartson, N. Dwight Barnette
Title: Interactive Learning with a Digital Library in Computer Science

C.1.2. Summary of Publications

Project Overviews: FOXE94e, FOXE95c, FOXE96d, FOXE96f
Digital Library - General: CHEN96, DOUG95, FOXE93b, FOXE93d, FOXE93e, FOXE94d, FOXE95a, FOXE96b
Digital Library - Architecture: FOXE94b, FOXE95f, GLAD94a, GLAD94b
Digital Library - Capture, Conversion: DALA93
Digital Library - CS Technical Reports: FOXE95b, FREN95, MALY94a, MALY94b
Digital Library - Education: FOXE96d, LAUG96
Digital Library - Interface: NOWE94, NOWE96, WAKE95a
Digital Library - System: FOXE93c, FOXE94c, HEAT95
Interactive Applications: WAKE95b, WAKE95c
Interactive Learning: SHAF96a, SHAF96b, TINO96, YANG95
Networking: ABRA95a, ABRA95b, FOXE94a, WILL96

C.2. Problem, Purpose, Goals, Objectives

Problem: There is a drastic shortage of people with suitable educational background to work in the broad field of "Multimedia, Hypertext and Information Access" (MHIA). There are no guidelines available for curricula in this field. Few universities have courses in the area, and those courses suffer from a severe lack of suitable textbooks. New instructors often start from scratch to construct course notes or to design laboratories that are important for hands-on experience with tools or to support group projects. There are no comprehensive readings volumes or bibliographies. There are few collections of case studies or projects, few sets of exercises, and no large test banks. Popular interest and curiosity has created a tremendous demand for knowledge that is largely unfulfilled, and many misconceptions or limited understandings have resulted:

- People are not aware of research in electronic publishing, the aims of SGML, or its relationship to HTML.
- People are not aware of work in the hypertext field that offers solutions to many present difficulties faced by developers and users of the WWW.
- People exploring digital libraries are unaware of decades of research in information retrieval or library and information science that can make searching in the WWW more efficient and effective.
- Developers of multimedia applications are unaware of key concepts and tools prepared by those researching into authoring systems, that would provide a framework and support for their efforts.

While many of these problems could be solved by curriculum development efforts, others require closer connections between the research and education groups working in the field.

Purpose: This project aims to solve the above-mentioned problems by applying research in the MHIA field to. enhance education in that field --- both by making research results available and by using that research to make education more effective.

Goals: As a result of this project:

- Universities should be able to add or enhance modules to existing courses to introduce concepts from the MHIA field.
- Universities should be able to add or enhance undergraduate and graduate level courses in or related to MHIA.
- Universities should be able to add or enhance curricula and programs in the MHIA area.
- Educational activities in the MHIA field should be able to draw upon a rich and continuing "feed" of research results, to increase motivation, and ensure timely updating of content (in this rapidly evolving area).

Objectives: The key objectives relate to curricula, research, digital library development, and evaluation:

- Curriculum will be developed for MHIA in accordance with the two leading computer professional societies: ACM and IEEE Computer Society. (See letters in Section I from ACM, the ACM Special Interest Groups on hypertext (SIGLINK), information retrieval (SIGIR), and multimedia (SIGMM) as well as the IEEE-CS Multimedia Computing Technical Committee.) We will help develop a unified curriculum during the term of this project, to be co-published by all of these groups and widely disseminated. We also will work with other engineering groups around the nation that are interested in curriculum to help them meet the rapidly growing need for trained professionals in the MHIA area.
- A digital library (courseware repository with suitable cataloging, linking, searching, annotation) will be created for the MHIA field, to include: data collections for all media forms (text, image, audio, video); software packages (with documentation and test suites); demonstrations (interactive when possible, else slide shows or scripted sequences of screen dumps); simulations; bibliographies, articles, technical reports, pointers to theses/dissertations, and other publications; laboratory session documentation; case studies, real-life project reports; syllabi, course calendars, lecture notes, tutorials; exercises, assignments, problem sets, quiz/test banks; animations (including of algorithms and processes), multimedia presentations; and Web based training, computer aided instruction.
- Mechanisms will be put in place for the project staff to "prime" the digital library with educational and research materials, starting with their own works, and extending to include others' works they are aware of and rate highly. These mechanisms will make it easy for project institutions to continue to add to the repository even after the grant period is over. Included will be schemes for multi-institution collaboration, assessment, and decision making, as well as multi-hosted distance education.
- Collection development mechanisms will be put in place for the digital library to expand in a self-sustaining fashion, as part of the normal efforts of the professional societies --- a novel approach to coupling associations with education. All conference proceedings (from IEEE-CS events and those of ACM and its relevant SIGs) and journals of the societies may be made available electronically to project participants, with suitable intellectual property rights management policies enforced. All presenters of demonstrations, videotapes, tutorials and other conference-related activities will be asked to contribute their materials to the emerging repository. Submission specifications will be given to authors upon acceptance of their works, and project staff will check and catalog each entry. Project staff will collect and analyze ratings of submissions, and attach recommendations regarding educational utility, ensuring easy incorporation into curriculum.
- Evaluation mechanisms will be put in place so that logs, usage records, remote evaluation data, user ratings and comments regarding the digital library and its content will be collected, analyzed, and made available. Evaluation will encompass formative efforts to improve the quality of "documents" in the digital library, as well as summative studies of their value. Special mechanisms will be provided so that those interested in coop, intern, and hiring programs can provide criteria, direction, and post-graduation data to guide the evaluation.

C.3. Participants

C.3.1. ACM and IEEE-CS

ACM and IEEE Computer Society have long been the key players in curriculum development for the computing field [ACMC68, ACMC77, ACMC79, ACMI91, ACMS92, DENN88, IEEE76, IEEE83]. From the support letters in Section I (ACM, SIGIR, SIGLINK, SIGMM, IEEE CS TC on Multimedia Computing) it is clear that this

project has support from interested parties in those societies. Further, close working relationships between those groups and project staff already exist. ...

C.3.2. Five Universities

Carnegie Mellon University (CMU), Cornell University, Texas A&M University (TAMU), University of California at Berkeley (UCB), and Virginia Tech (VT) have played a lead role in many advanced research efforts in this field. ...

C.3.4. International Connections

The MHIA field has long had international aspects, though the U.S. does lead in the research arena. In Information Retrieval there are strong groups in Australia, Denmark, France, Germany, Italy, U.K, and other locations. Digital Libraries work has spread to Australia, Germany, Korea, Singapore, U.K. and other nations. In the Hypertext field there are strong groups in such countries as Italy and U.K. In the Multimedia arena, large groups are working in places like France, Japan, Singapore, and U.K.

Because of the growth of the industry and scarcity of trained people in the MHIA field, the European Community has provided funding to develop a syllabus for an M.Sc. in what they call "Information Engineering." That effort involves industry, as well as educators from Dublin, Sheffield, Padua, Balleiric Islands, Glasgow, and possibly other locations. The M.Sc. will involve a common "course", common exams, and common delivery using video conferencing and the Internet. The letter in Section I from Pantry Associates, which coordinates the effort, indicates willingness for that group to work with this project's staff. That will be of particular importance for broad acceptance of recommendations since the international research community is so tightly coupled in this field. One might argue that the EUROIEMASTERS Project demonstrates European cost sharing with this project --- the concept paper version of this proposal was used in an August 1996 meeting in Mallorca in which initial curricula and program plans for Information Engineering were crafted.

C.4. Related Work

C.4.1. Workshops

Work toward curriculum development for MHIA has proceeded since 1993, according to the following calendar of workshops and other events. PI Fox has been involved in all these activities.

Year	*Activity*
1993	ACM SIGIR Education Committee appointed
1993	NSF Information Engineering Task Force had initial meeting Nov. 4-6, Alexandria, VA
1995	Panel on Education in Information Retrieval at ACM SIGIR'95
1995	Curriculum Development in Computer Information Science: A Framework for Developing
	a New Curriculum in IR, workshop following ACM SIGIR'95
1995	ACM Multimedia'95 business meeting discussed education needs, and led to appointment
	of Education Committee co-chairs
1996	Information Retrieval 2000: Workplace Needs & Curricular Implications (Drexel, Kellogg)
1996	Two Birds of a Feather discussions on multimedia education at ED-MEDIA'96
1996	Courseware, Training and Curriculum in Information Retrieval workshop held after SIGIR'96
1996	EUROIEMASTERS Project meeting in Mallorca, for Information Engineering Program
1996	Courseware, Training and Curriculum in Multimedia workshop, planned for Multimedia'96

Table 2. Key MHIA Curriculum-Related Events

C.4.2 SIGCHI Efforts

ACM SIGCHI commissioned a group to work on curricula in the human-computer interaction field, which led to very influential guidelines [ACMS92]. Those serve as a model for parts of this proposed effort. Since our project

will adapt that model, it is appropriate to provide some details. See Table 3, based on the Table of Contents of [ACMS92], which shows the various sections of the curricular document.

It is noteworthy that there are a number of types of overlap between HCI and MHIA issues. Thus, first, the human sensory systems (e.g., human visual and auditory systems) are of interest to both fields. Second, developing interactive applications is a challenge in both disciplines. Third, multimedia concepts of quality of service relate directly to HCI issues of usability. Fourth, both fields have dimensions that fit into computer science, management information systems, and interdisciplinary programs. Finally, each relates to a number of areas of computer science (e.g., video or image processing, software engineering, graphics, databases, data structures, audio processing, algorithms, AI) and so can be taught, in part, in connection with those courses.

As a result, in connection with the various events listed in Table 2, PI Fox was contacted by a number of experts in the HCI field, asking if they could attend, since they were developing new programs in the multimedia area. One such example is an effort involving HCI researchers (e.g., Ron Baecker) at the University of Toronto Knowledge Media Design Institute.

Section	*Contents*
Chapter 1	Introduction
Chapter 2	Human-Computer Interaction
	2.1 Definition of HCI
	2.2 Field of HCI
	2.3 The Content of HCI
Chapter 3	Courses in HCI
	3.1 The Four Proposed Courses
	3.2 Sequencing Courses
	3.3 Structure of the Course Descriptions
	3.4 Course Descriptions
Chapter 4	HCI Curriculum Designs
	4.1 HCI-oriented, not HCI-centered Programs
	4.2 Base Disciplines for an HCI Orientation
	4.3 Adapting Existing Programs to an HCI Orientation
	4.4 Framework for a Program Based in Computer Science
	4.5 Framework for a Management Information Systems Program
	4.6 Interdisciplinary Programs for HCI
Chapter 5	Issues Raised by our Recommendations
	5.1 Implementation Suggestions
	5.2 Unresolved Issues
	5.3 Known Limitations
	5.4 Conclusion
References	
Appendix A	Resources for HCI
Appendix B	An Information Systems Curriculum in HCI
Appendix C	A CS Undergraduate Specialization in HCI
Appendix D	An Interdisciplinary Undergraduate Specialization in HCI
Appendix E	Example Course Taught in HCI
Appendix F	Case Studies in HCI

Table 3. Adapted, Abbreviated Table of Contents for "Curricula for Human-Computer Interaction"

C.5. Research and Curriculum Development

This project deals with two closely coupled types of effort. One relates to curriculum development, and the second focuses on connecting research with education. Accordingly, these two matters are explored in the next two subsections, followed by a discussion of how the two efforts can be integrated.

C.5.1. Curriculum Development Overview

Based on the results of discussions at previous workshops, we may classify courses in the MHIA area using the following taxonomy:

A. Level:
{Introductory, Advanced} x {Undergraduate, Masters, Ph.D.}
B. Disciplinary Focus:
{Arts, Computer Science, Computer Engineering, Education,
Information Systems, Library and Information Science, }
C. Depth:
{1-credit overview, 3-credit course, 6+ credit sequence, concentration, major}
D. Pedagogical Style:
{Lecture, Lab-based, Self-study, Case-study, Group project, Seminar}
E. Approach:
{Theory, Design, Application Development, Tool Use}

Discussion below shows how this might relate to observed needs, and be a manageable problem in terms of curricular development.

C.5.1.1. Relevance

As mentioned earlier, the MHIA area has become very popular, with serious shortages in the number of trained personnel. The range of jobs, however, spans from technical ones (dealing with development of hardware, software, and application development tools) to user-oriented ones (in which one learns how to use tools) to service-oriented ones (in which some application is developed or a service rendered). Preparing people for those jobs thus requires varying degrees of Level, Depth and Approach. If a curriculum is comprehensive in covering all the options, then it is guaranteed to be relevant to the identified needs.

C.5.1.2. Innovative Approaches

Because of the complex taxonomy in C.5.1 above, the fact that MHIA covers the domain of at least three ACM Special Interest Groups, and the rapid evolution of new technology in this area, special approaches are needed if this project is to achieve adequate coverage. Accordingly, we introduce a number of innovative approaches:

- Adopt the re-use approach. If courseware can be developed in small, modular chunks (which is encouraged by use of WWW pages), then it can be used by the same instructor in other courses, other instructors, or even people in other institutions. Re-use can work with an "approximate match" instead of the exact fit that makes software re-use so difficult.
- Adopt a 3-level model. Rather than function only at the program and course levels, also have a lower level of "knowledge modules". If these are small enough, they can be self-contained and can be more easily re-used. If developed for self-study, they can be learned or reviewed when needed, fitting well into the requirements of life-long learning.
- Apply concepts of technology transfer to connect research with education. Technology transfer is a difficult problem, depending on proper organization structure, long-term relationships between groups that play differing roles, and building suitable bridges [ISAA96]. Prototypes sometimes help ensure effective communication between those groups [SCHO96]. These need to fit into the broader context of a "content life cycle" if they will affect business opportunities [KUCH96]. Especially regarding industry-university collaboration, a great deal of contact and interaction is needed [FOLE96]. All these points indicate that students should: be exposed to prototypes and demonstrations, have frequent contact with researchers and each other (especially in project teams), and that a good social support structure is very important for learning.
- Leverage activities of professional societies to afford sustainable enhancement to curricula. Since conferences in the MHIA area often have demonstrations, or presentations of prototypes, or videotape overviews, we propose to work closely with conference groups so that good demonstrations be "canned" and made available to other faculty and students.
- Use digital libraries to implement the old notion of courseware repository. We propose to build upon our various digital library activities, and to extend the notion of sharing facilitated with the WWW, by having a digital library of MHIA content, to include especially demonstrations, relevant publications, test data, multimedia objects, programs, animations, etc. These will be carefully cataloged using appropriate metadata, and be safeguarded with intellectual property right sensitive access tools.

- Use the technology studied to learn about it. Since multimedia content is often motivational, or provides alternate ways of access so that people with different preferences can be accomodated better, it will be used as much as possible to help learning. Since hypertext provides user control, flexibility and ready re-use, it also will be applied. Students will learn both about MHIA and about how to apply or use it.

C.5.1.3. Example Curriculum and Courses

Taken from the Computer Science domain, the following illustrate typical courses now offered at institutions involved in the project.

Virginia Tech's **CS4624** is a senior level course on Multimedia, Hypertext and Information Access:
CS4624 Catalog Listing: Introduces the architectures, concepts, data, hardware, methods, models, software, standards, structures, technologies, and issues involved with: multimedia information and systems; hypertext and hypermedia; networked information; electronic publishing; virtual reality; and information access. Students will learn how to capture, represent, store, compress, manipulate, interact with, and present text, drawings, still images, animations, audio and video. They will work with video conferencing, authoring systems, and digital libraries.
CS4624 Additional Description: This course is designed for seniors to become familiar with a range of information technologies, much like Database Systems covers certain types of data, and Artficial Intelligence covers certain types of knowledge. Coverage includes text, electronic publishing, search, retrieval, browsing and related issues of information access. Other media types are considered separately and in combination, when synchronization and time-based performance are crucial to achieve adequate quality of service. Linking, hierarchical structures, streams, layers, and similar organizations and views will be considered. A wide variety of hardware, software and application demonstrations will be given. Students will gain expertise working with key packages such as AuthorWare, Mosaic/Netscape, and Storyspace.

Cornell University's **CS631** is graduate course that discusses enabling technologies and advanced applications in multimedia. Topics include analog image, audio, and video, sampling theory, color and perception, compression techniques for image, audio, and video data (JPEG, Musicam, MPEG, and H.261), compressed domain processing, image and video processing (morphing, mosaicing, object tracking, image comparison, cut detection), networking for multimedia (multicast, guarenteed quality of service), video storage systems (admission control, disk buffering and scheduling, and storage hierarchies), operating system scheduling, and toolkits for multimedia.

Virginia Tech's **CS5604** is an introductory graduate level course on Information Storage and Retrieval.
CS5604 Catalog Listing: Analyzing, indexing, representing, storing, searching, retrieving, and presenting desired information. Models, document processing, thesauri, evaluation of system effectiveness, special hardware. Boolean logic and inverted file systems. Fully automatic systems. Role of probability, artificial intelligence and computational linguistics.
CS5604 Additional Description: Explanation/demonstration of: online bibliographic services; library systems like VTLS, MARIAN; retrieval systems like WAIS, CODER; hypertext systems like WWW, Hyper-G, KMS; digital multimedia with JPEG, MPEG; applications of artificial intelligence in knowledge bases and information systems; text processing, electronic publishing, automatic indexing. Course theme: "digital libraries."

C.5.1.4. Integration with Curriculum

MHIA curricula must fit in with various currently existing curricula, particularly those related to computer science and engineering. In some cases, the approach will be to add a small number of knowledge modules to existing courses, such as:

- architecture,
- artificial intelligence,
- computer architecture,
- data structures,
- database management,
- graphics,
- human-computer interaction,
- networking,
- operating systems, and
- programming languages (e.g., scripting).

Alternatively, using our 3-level model, student may take a course or a higher level collection (sequence, concentration, minor, or major). ...