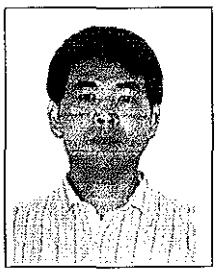
Connecting Educational Information Spaces

Roy Rada*, Zhengije Liu**, and Min Zheng***

* Washington State Universita, Pullmann, Washington
** Dalian Maritime University, Dalian, China
*** Information Technology Institute, Singapore



Author Biographies:

Roy Rada is the Boeing Distinguished Professor of Software Engineering at Washington State University. Immediately prior to that he was a Chair of Computer Science at the University of Liverpool, and before that the Editor of Index Medicus at the National Library of Medicine in Bethesda, Maryland. He holds B.A., M.Sc., M.D., and Ph.D. degrees.

Zhengjie Liu is an Associate Professor of Information Systems at Dalian Maritime University. Previously, he was a visiting scientist in Liverpool.

Min Zheng is a senior member of technical staff for electronic publishing at Information Technology Institute in Signapore. Prior to that he was a visiting scientist at the University of Liverpool and before that an associate professor at Peking University where he earned his Ph.D. in Computer Science.

Rada, R., Liu, Z., & Zheng, M. (1997). Connecting educational informational spaces. *Knowledge Organization, 24(3),* 154-162. 20 refs.

Abstract: Online textbooks can be connected to other sources of information to improve their educational value. We describe two case studies. One used a medical textbook and connected to medical journal abstracts via a thesaurus. The textbook, journal abstracts, and thesaurus were stored on a CD-ROM. The other case study shows a textbook on the World Wide Web that is connected to various other sources of information. About half the book references are to web sites, and the textbook is part of an online course that is connected to an online catalog and other courses. Such linkages among information spaces should help students navigate the information relevant to their studies.

1. Introduction

The rapid growth of information in the world continually challenges us to find new tools to help us discover and appreciate the knowledge behind the information. Educators and researchers have organized this information in various ways. For the purposes of this work, a textbook is called a primary information source for a student. A database of journal article abstracts is an example of a secondary information source. What is the significance of this distinction for learning and for building computer-based systems to help learning?

The process of learning may be an active and exploratory one in which the learner decides what and how to learn (Boud, 1990). The human learning process requires searching for, selecting from, and reorganizing of information (Howes and Payne, 1990). Courseware should help the learner construct learning opportunities (Reisman and Carr, 1991) through learning by doing (Schank, 1994), and should help the learner to build a cognitive map of the learned knowledge (Geyer, 1986) which links to the existing knowledge framework.

When a learner encounters a concept and tries to understand it, he/she may want to compare it with the definition of another concept previously learned or to consult a relevant paper about a specific example. Related knowledge sometimes comes from textbook materials (primary information), but sometimes also involves accessing the secondary information space. Because there is a gap between the two information spaces, the links which bridge the gap may support the learning process (Rada and Murphy, 1992) (see Figure 1). In non-computerized learning environments, this kind of link may be provided in

paper-based form and the corresponding information access must be done manually (Nowaczyk and Synder, 1993).

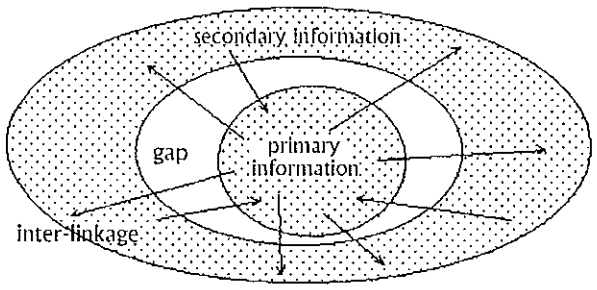


Figure 1. Primary information space and secondary information space.

Courseware that provides access to both primary information and secondary information requires large-volume storage. This storage might be provided by networks as through the World Wide Web or on CD-ROMs. This paper describes in detail a project that converted a medical textbook and secondary information source into CD-ROM-based, interactive, hypermedia courseware (Peng et al, 1994) and then describes more briefly a web-based courseware project.

Our work is that of engineers. We have built systems that have novel features and that demonstrate important architectural features for connecting textbooks and secondary information spaces. We do not present extensive reviews of alternative models of information structuring (e.g. Agosti et al, 1996). We do not present experimental results about the utility of our system in education. We readily acknowledge that in the end the value of the engineering must be demonstrated by its usability in a systematic way (Draper et al, 1996).

2. Medical CD-ROM

We were partners with Elsevier Scientific Publishers in two consecutive European Union projects, one called OAR and the other GALEN. In the OAR project we developed algorithms for connecting documents to queries via intelligent processing of the thesaurus for Excerpta Medica (Rada et al, 1991). In the GALEN project we developed new medical thesauri to connect patient records and the medical literature (Rada et al, 1992). We were also working with funding from the British government in a consortium of medical schools to develop medical courseware. One of the partners in this medical school consortium had co-authored a popular medical textbook entitled *A Colour Atlas and Text of Clinical Medicine* (Forbes and Jackson, 1993). Forbes was willing to work with us to develop an interactive, electronic course based on his book. The publishers of the Forbes and Jackson textbook were eager to pursue electronic publishing and contracted with our group for the development of a prototype CD-ROM to satisfy several properties. We developed two prototypes of our CD-ROM and believe that aspects of our prototypes are of interest to the academic community.

2.1 Medical Information Spaces

The design of the system must start from an analysis of the components in the information space, as well as their relationships. For our purposes, the information space for the learning of clinical medicine is composed of the following information components:

- the clinical medicine textbook,
- a set of journal abstracts of medical literature, and
- a medical thesaurus.

The 'Colour Atlas and Text of Clinical Medicine' (Forbes & Jackson, 1993) (hereafter abbreviated Clinical Medicine Book) constitutes the primary information space. The readership would include medical students and candidates for postgraduate examinations, while also being of use to a wider medical audience in other clinical specialties. Each chapter contains extensive cross-references (intra-links) to aid the effective use of the book and its illustrations. There are a total of 1434 images, which include full color photographs, microscopic, radiological, ultrasound and endoscopic images. The index includes the key terms which appeared in the textbook as well as their location in the textbook.

EMBASE is the Excerpta Medica online database, which contains biomedical information taken from 4500 journals published in many countries and languages (Excerpta Medica, 1989). EMTREE is the numbered classification system used to index EMBASE. The EMTREE system is based on fifteen facets ranging from Anatomical Concepts to Society and Environment. Each facet contains up to twelve levels of hierarchy.

The Excerpta Medica Abstract Journals are derived from EMBASE. Medical specialists select the most important abstracts from EMBASE to be printed in the abstract journals. The Excerpta Medica Abstract Journals include 45 sections which together cover the complete field of biomedical science. Excerpta Medica Abstract Journals are published several issues every year, each of which is dedicated to the topic of one of the 45 sections. Abstracts in each issue are classified into several categories. By providing access to the Excerpta Medica Abstract Journals, the Clinical Medicine Book would be further enhanced by thousands of references to specific, related information in the area of clinical medicine.

The information access across the two spaces can be classified into three types:

- The request of information access is generated in the primary information space and satisfied in the secondary information space through index/thesaurus.
- The request of information access is generated in the secondary information space and satisfied in the primary information space through the index/thesaurus.
- The information access request is generated either in the primary information space or in the secondary information space, and satisfied by browsing the thesaurus.

The composition of the information space and the relationships among the components highlights the importance of the thesaurus and index (see Figure 2).

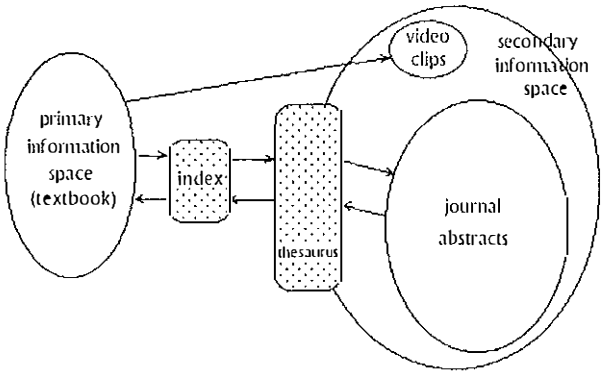


Figure 2. The primary information space and the secondary information space connected by index/thesaurus. Video clips were also incorporated into the courseware.

2.2 *Prototype Design and Implementation*

A prototype has been designed and implemented. The purpose of a prototype is to refine and test the models earlier forwarded. The prototype is built with a hypermedia authoring tool called Guide. The storage medium was a CD-ROM.

While EMTREE covers almost every specific topic in medicine, the Clinical Medicine Book covers only the major topics. Many parts of EMTREE were of no help in providing further information on topics in the textbook. How shall one combine useful parts from EMTREE and the Abstract Journals?

The starting point in the selection of related material from EMTREE and the Abstract Journals is the index of the Clinical Medicine Book. Both the textbook index and EMTREE were made available in electronic form. With the online index and EMTREE, one could use the computer to find the terms in both. The EMTREE terms which also appear in the textbook index are considered important. These terms were marked for further processing. After finding the related terms, related journal abstracts could be found based on these terms. In the Abstract Journals, each section includes a subject index which is an alphabetic list of all the abstracts in this section based on the keywords. If one of the keywords of an abstract falls in the category of a related index term, these abstracts were marked as a related abstract and included in the CD.

The index of the textbook and the medical thesaurus (Sub-EMTREE) constitute a bridge between the primary information space and the secondary information space. The index is from the textbook itself, so each term has at least one link to the corresponding term in the body of the textbook. The thesaurus is a subset of EMTREE with each branch including terms appearing in the index. These terms are treated the same as the terms in the index, i.e. linking to the corresponding term in the body of the textbook. The links between the textbook and the index/thesaurus are made bi-directional so that one can go from a term in the body to the corresponding position in the index or thesaurus to view other appearances of the same term, if any, or to check related concepts through the thesaurus. To make a reverse link from text to index, the natural way is to mark the term in the textbook as the start point of a reference link and the end point is in either index or thesaurus. When a term is linked from both index and thesaurus, the reverse link is linked to the thesaurus because a term in the thesaurus includes all links associated with a term in the index. With these reverse links, one is also able to check related material through the thesaurus.

At the same time, the index and thesaurus are also linked to the journal abstracts. Since the abstracts are selected based on the index terms of the textbook, each selected abstract should be related to at least one index term in the textbook. Therefore, links can be established between the selected abstracts and the index, as well as the thesaurus. With these links, one can quickly go to related literature abstracts from the textbook through the index or thesaurus.

2.3 *Examples*

The following examples show the three types of information access:

- from the primary information space to the secondary information space,
- information access in the reverse direction, and
- a thesaurus providing a unique perspective on the information structure.

When a learner encounters a concept in the primary information space, namely the textbook, s/he sometimes wants to find more knowledge even though at the moment s/he does not know clearly what is needed. S/he may go to and browse the the-

saurus, find some interesting terms, and then go to the corresponding journal abstracts for more specific knowledge. For example, the learner encounters the concept `liver infection' in the textbook and needs to know more about it. S/he may then jump to the thesaurus through the textbook-thesaurus link. In the thesaurus s/he finds related concepts (see Figure 3), selects 'hepatitis', and goes to the corresponding journal abstracts entitled:

- "Human immunodeficiency virus as a possible cofactor in the development of fulminant hepatitis B in intravenous drug abusers"
- "Summary of the biosafety workshop: application of biosafety principles in blood establishments"

by the thesaurus-abstract links, from which more specific knowledge can be obtained.

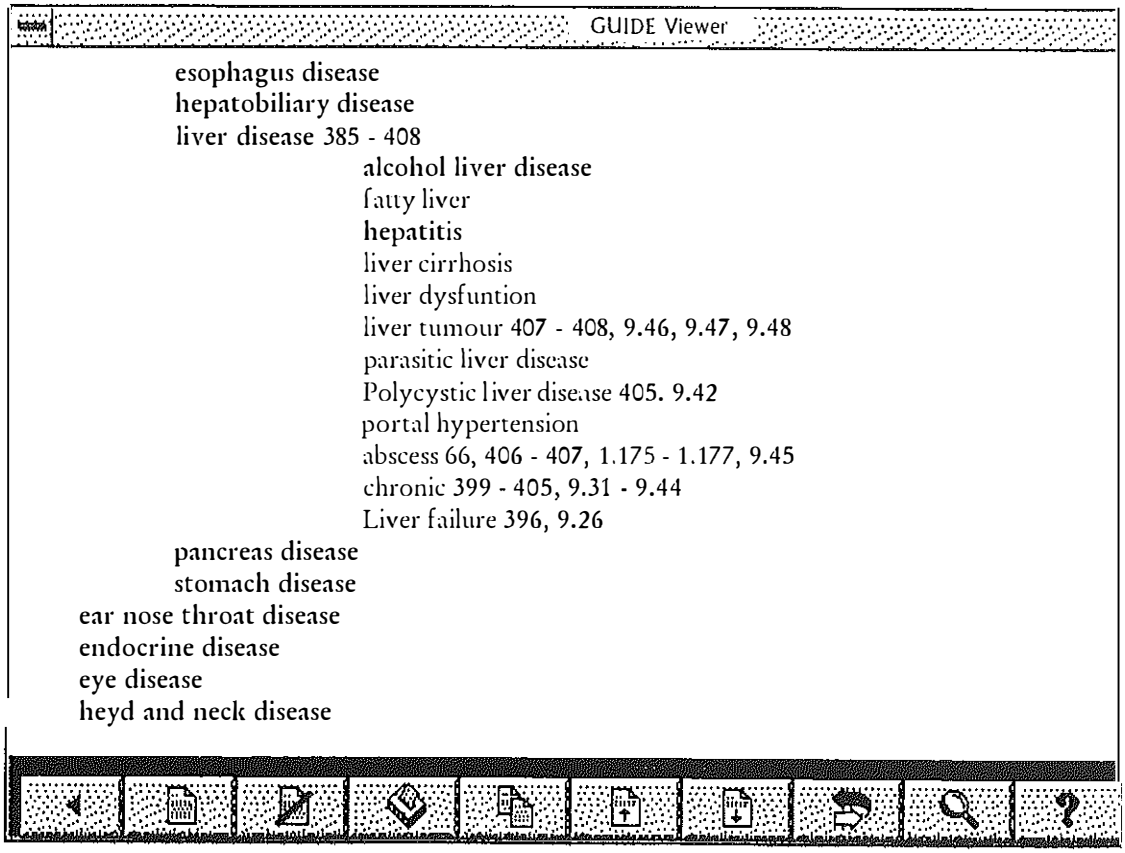


Figure 3. Screen dump showing the part of the thesaurus related to `liver disease' from where one can further jump to the abstracts related to `hepatitis'

Textbooks usually emphasize the systematic introduction of knowledge, while journal papers emphasize research results. These two information sources can complement one another. For example, when the learner reads the journal abstract entitled "Hepatitis ◗ virus and human immunodeficiency virus antibodies in parental drug abusers who are hepatitis B surface antigen positive", s/he might want to know the basic properties of hepatitis ◗ and B which are not available in the abstract. In this case, through the thesaurus, s/he can go to the corresponding paragraphs of the textbook to get a systematic introduction to them.

The structure of the textbook which is outlined by the table of contents of the book provides one perspective, while the thesaurus provides another. When a learner reads the textbook, s/he sometimes feels that a concept needs to be understood in a broader perspective. Through browsing the thesaurus, the learner can get a different insight into the concept by being aware of the relationships among related concepts, so that a deeper understanding can be formed. And also on this basis, s/he can go to the relevant part of the textbook or corresponding journal abstracts to get more information. For example, the learner encounters the concept `portal hypertension' in Chapter 9 of the textbook and might go to the textbook table of contents (see Figure 4a). By browsing the thesaurus (see Figure 4b), s/he can get a different perspective.

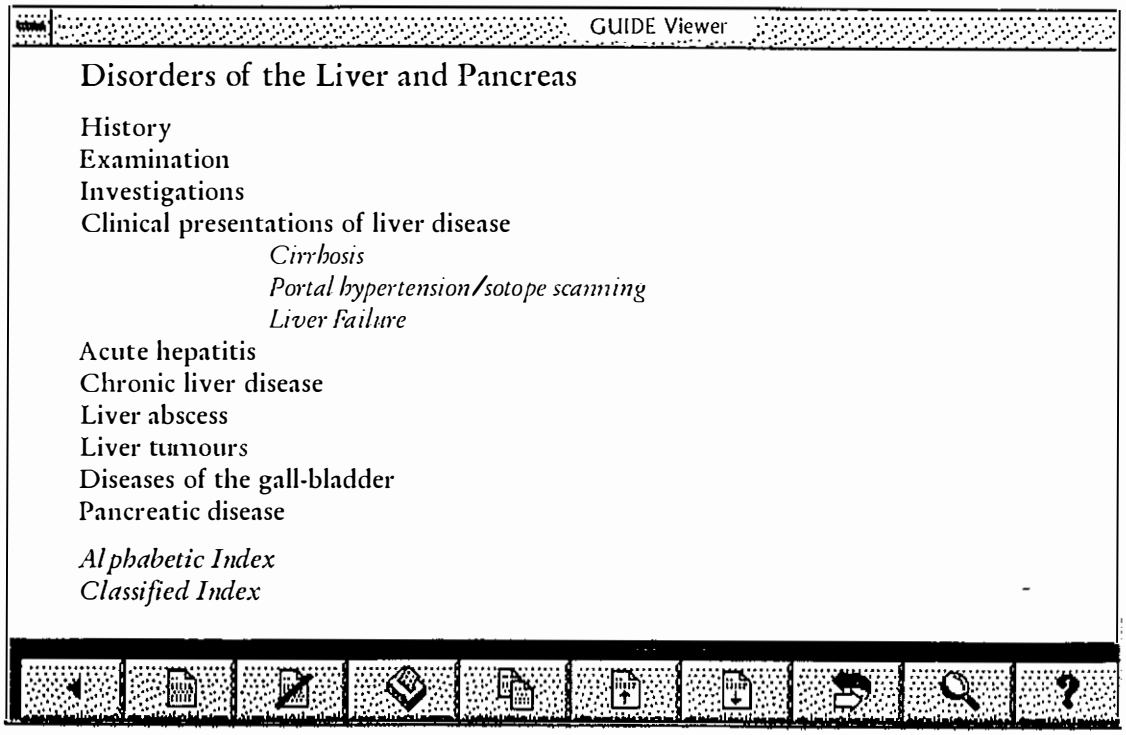


Figure 4a. The heading `portal hypertension' can be viewed from the perspective provided by the textbook.

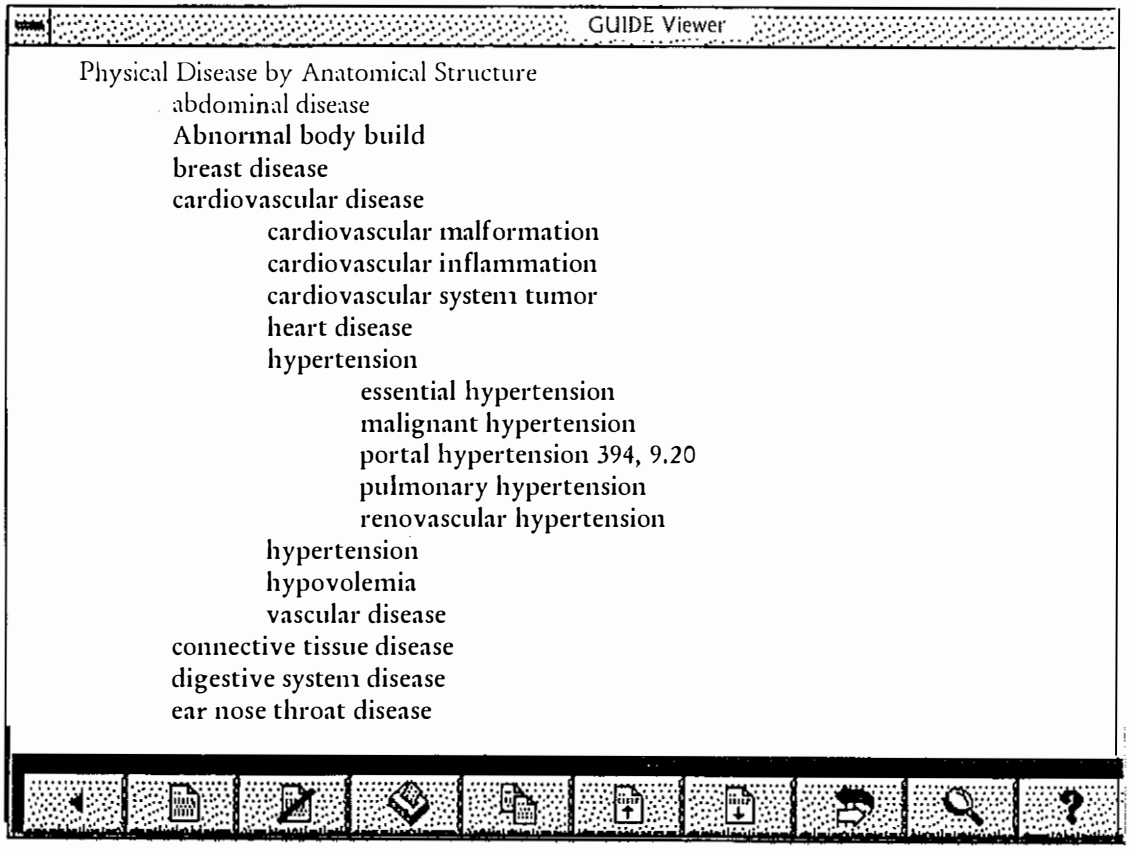


Figure 4b. The heading 'portal hypertension' can be viewed from the perspective provided by the thesaurus (which is different from that provided by the textbook).

We have developed connections between a medical textbook and Excerpta Medica in a semi-manual fashion. Terms in portions of the textbook were semi-automatically collected and matched against terms in Excerpta Medica. Some matches were then chosen in an intuitive fashion to incorporate in the prototype electronic product. Our intention was to demonstrate in subsequent prototypes how these processes could be extended both with automatic tools and with systematic user feedback. The automatic tools would include more sophisticated programs for selecting terms from the textbook and more sophisticated algorithms for matching these terms against components of Excerpta Medica. Advances in the methodology would include allowing the user to define a profile that would determine which matches between the textbook and Excerpta Medica were made. To effect dynamic matches to Excerpta Medica based on user profiles would require that large amounts of Excerpta Medica be accessible. One way to do this is to have the CD-ROM invoke access to Excerpta Medica across computer networks. The CD-ROM is a particularly appropriate medium for videotapes and large amounts of static textual information. Software on the CD-ROM can also initiate communication with the World Wide Web and bring many options to the user.

3. Web Textbook

The notion of primary and secondary information space is only a first approximation to separation of information spaces. Information may be categorized along multiple facets and along each facet it might have many attributes. The first author has developed another hyperlinked textbook which illustrates further aspects of connecting across information spaces. The book is available in paper and online as part of a course, and the course is part of a virtual college.

3.1 Virtual Education Manifesto Textbook

The textbook is entitled *Virtual Education Manifesto* (Rada, 1997a). The paper version of the book provides the traditional

1. outline to connect components of the book,
2. an index of alphabetically sorted terms that point to relevant components of the book, and
3. bibliographic references that point to secondary sources of information.

The textbook includes 175 references. 81 are references to material on the web. In the online version of the textbook, these links are active so that the reader can select the reference and be then taken to that reference across the web. These "bibliographic" references vary widely in character. Some are to preprints of journal articles. Some are to papers that were presented at workshops (see Figure 5). Some links are to

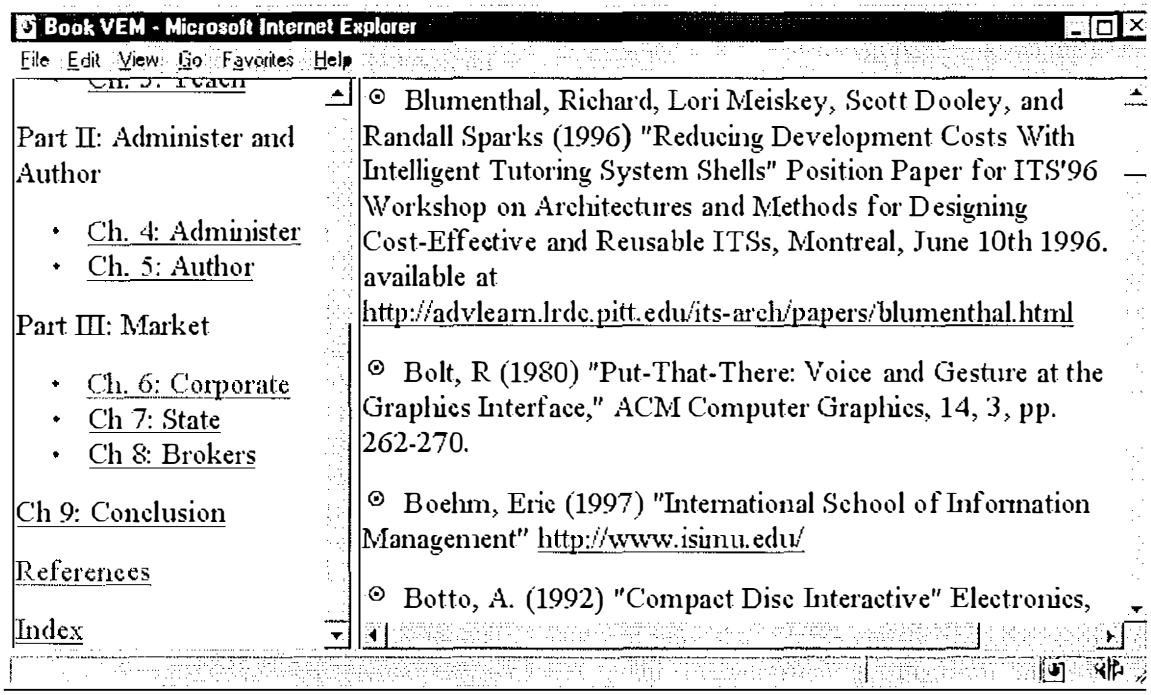


Figure 5: Screen from online *Virtual Education Manifesto*. The table of contents is on the left. The References section has been selected and shown on the right are actual references. Notice here that 2 of the 4 references are to web material -- a ratio similar to that throughout the References section.

web sites of institutions that the textbook is summarizing. Some links are to other online textbooks. For topics like virtual education, the most important references are often published on the web and nowhere else. For online textbooks in certain domains, one can expect to find an increasing number of references to web-based material.

The online version of the *Virtual Education Manifesto* is part of a course being offered in the Virtual Information Technology College of the Globewide Network Academy. The Globewide Network Academy is itself a massive catalog of distance education courses (GNA, 1997). Each course is characterized with about twenty attributes. One of these attributes is the topic of the course. The topic must be described with some set of terms from a classification that GNA provides - a classification of topics for distance education. The course entitled „Virtual Education" is currently classified under three concepts of education, computers, and Internet. By browsing or searching through the catalog to these topics one comes to a listing of the Virtual Education course. The course listing (see Figure 6) provides values for attributes that include instructor email address, method of evaluation of students in the course, format of delivery of material, cost of enrollment, dates of course offering, and other such material about the actual offering of the course.

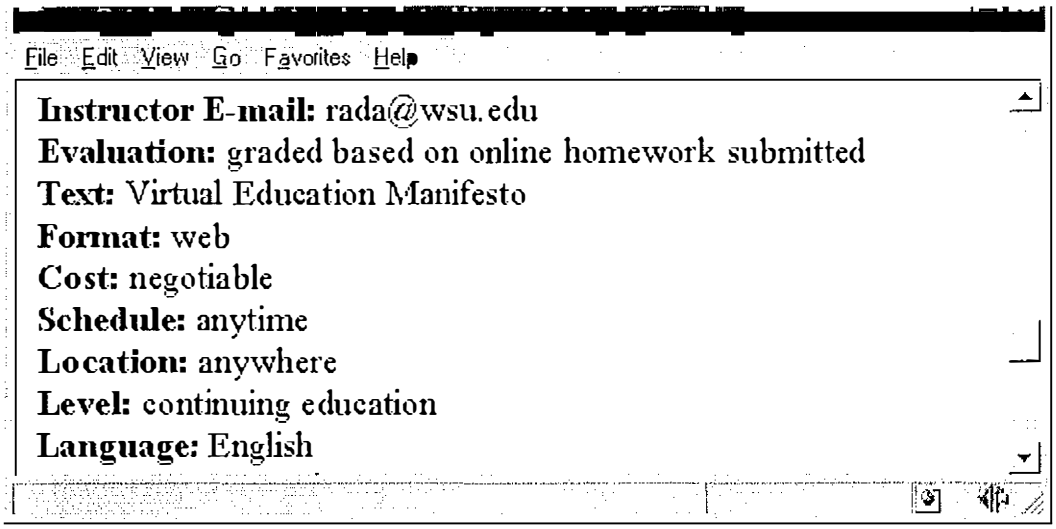


Figure 6: Screen from GNA Catalog entry for Virtual Education course. A few of the attributes are shown on the left hand side in bold and the attribute values are indicated on the left.

The catalog also carries information about the prerequisite courses, if any, for the course in question. This kind of attribute can be an active link, given that the other courses are also available on the web. In this sense the course is primary information and is connected to other courses in the catalog as secondary information.

3.2 Standards to Facilitate Connectivity

To facilitate connectivity of educational material from any given primary source to related secondary sources a variety of standards have been developed. These standards are not necessarily widely followed but they each have an audience and are expected to become increasingly important (Rada and Schoening, 1997b). We look briefly next at two standards as they are germane to the theme of this paper. One is a standard from the Aviation Industry Computer-Based Training Committee intended for courses and classrooms. The other is a standard from Educom that targets connectivity across the web from institution to institution.

The Aviation Industry Computer-Based Training Committee (AICC) standards specify that the media components of a course must be stored in one of a few common formats. They also specify that the courseware must be associated with a text file that describes the course structure and specifies the learning objectives of each component of the course and how these learning objectives relate to one another (AICC, 1995). AICC has also standardized how a piece of courseware will communicate with a classroom management system (AICC, 1997). The courseware must be able to import descriptions of students who are enrolled in the course and must be able to export information to the classroom system about each student's performance in using the courseware. Thus these AICC standards describe how the components of a course hang together and how student information is passed to and from the course system.

Educom is a consortium of higher education institutions that is focused on the best use of information technology for higher education and is attempting to both standardize and develop that technology. Educom has produced a standard called the IMS Metadata Specification (Educom, 1997). This specification is directed toward describing learning resources that are accessible, or perhaps just catalogued, online. The IMS Metadata Specification consists of three primary parts: a dictionary of terms, a description of learning resource types, and a system for managing the Specification. The IMS Dictionary identifies the terms that constitute IMS Metadata. These are the terms that are used to label the learning resources. IMS Metadata is broken down into fields and corresponding values. All of the proposed available fields are defined in the dictionary and their values are enumerated. These fields include author, credits, interactivity level, learning level, objectives, platform, prerequisites, price code, user rights, and user support. The hope of Educom is that all higher education institutions will use this metadata format to characterize their online educational material. Educom is working with the World Wide Web Consortium in the introduction of metadata information into the specification of the World Wide Web itself so that this metadata standard is conformant with other efforts at standardizing metadata formats across all web content.

The Virtual Information Technology College of the Globewide Network Academy intends to follow the standards advanced by Educom and AICC. The hope is that extensive connection of primary and secondary information sources will be facilitated in this way. Accordingly, students should be helped in their efforts to find valuable educational paths through the curriculum.

4. Conclusion

We have presented two case studies, one in detail and one briefly. Both illustrate methods of connecting primary and secondary information sources in online textbooks. One case is for a medical textbook connected via a thesaurus to medical journal article abstracts. The other case is for a web-based textbook connected to other web sites.

By combining the medical thesaurus from a secondary information space and an index from a medical textbook, we were able to create the spine of a merged information space. Our user scenarios involved, at least, three kinds of important link traversal: one from the book to the secondary information space, another from the secondary information space to the book, and finally within the thesaurus plus index. Our three examples from the prototype demonstrated these three cases. Such access is not easily supported in non-computerized learning environments.

A model for learning that entails linking textbook indices and secondary information space thesauri has been transformed into a design and then an implementation of a hypermedia system. Others could connect their own textbooks with their own relevant secondary information spaces in the same way. The example provided used a certain textbook, thesauri, indexed secondary information source, and hypermedia tools but the principles would apply to any combination of comparable information sources and tools.

The popularity of the World Wide Web gives new importance to the ability to first distinguish and then properly connect primary and secondary information spaces. The second example presents a web-based online course that includes another textbook. This textbook is richly connected to other web-based information and is part of an online college. The online college constitutes another, very different kind of secondary information space for a textbook. We have illustrated how the information structures associated with online courses, virtual classrooms, and virtual schools create numerous opportunities for useful linkages between the textbook and the other information spaces. The standards that are being most actively advanced in the educational technology arena are focusing on supporting the links within and among information spaces.

Research on the subject of connecting primary and secondary information spaces can go in a variety of directions. One direction only briefly addressed in this paper concerns the methods for semi-automating the establishment of linkages. Another direction is the construction of interesting examples. The authors are both developing information spaces and working on ways to connect them. One current project is to build a virtual school, replete with courses, catalogs, student profiles, and other entities. Each of these entities is part of an online information space and the challenge of the school is to correctly connect these spaces and to actively guide the users through the curriculum.

References

Agosti, M., Crestani, F., & Melucci, M. (1996). Design and Implementation of a Tool for the Automatic Construction of Hypertexts for Information Retrieval. *Information Processing & Management, 32(4),* 459-476.

AICC (1995). Courseware Interchange Guidelines and Recommendations. *Aviation Industry CBT Committee Courseware Interchange, AGR 007, Version*

1.0 by Courseware Technology Subcommittee at http://www.aicc.org/ agr007.htm

AICC (1997). Computer Managed Instruction Guidelines and Recommendations. *Aviation Industry CBT Committee Computer Managed Instruction, AGR 006, Version 1.1* by CMI Subcommittee at http://www.aicc.org/agr006.htm

Boud, D. (1990). Design for Learning Systems. *Making Learning Systems Work.* B. Farmer (Ed.). New York: Nichols Publishing Company. 3-11.

Dillenbourg, Pierre & Self, John (1996). What if the Computer Doesn't Know the Answer? *Communications of the ACM* August 1996.

Draper, S. W., Brown, M, Henderson, F., & McAteer, E. (1996) Integrative Evaluation: an Emerging Role for Classroom Studies of CAL. *Computers & Education. 26 (1-3),* 17-32.

Educom (1997). IMS Metadata Executive Summary. *IMS Metadata,* http://www.imsproject.org/ metadata/MDexec.html

Excerpta Medica (1989). *EMBASE Guide to EMTREE and Indexing Systems. (Vol.1-2),* New York: Elsevier Science Publishers.

Forbes, C. D. & Jackson, W. F. (1993). *A Colour Atlas and Text of Clinical Medicine.,* London: Wolfe Publishing.

Geyer, T., (1986). Probing a Student Teacher's Cognitive Structure of Software. *Journal of Computer Aided Learning, 3,* 89-98.

GNA (1997). Globewide Network Academy. *http://www.gnacademy.org/*

Nowaczyk, R. H. & Synder, A. D. (1993). Using Hypertext for College Course Information. *Computers & Education, 21(4),* 273-279.

Peng, H., Zheng, M., Keegan, D. & Rada, R. (1994). Combining Textbook and Literature on CD. *Medical Multimedia.* R. Rada and C. Ghaoui (Eds.). Oxford, England: Intellect Books, 62-81.

Rada, R. Barlow, J., Potharst, J., Zanstra, P., & Bijstra, D. (1991). Document Ranking Using an Enriched Thesaurus. *Journal of Documentation, 47(3).* 230-253.

Rada, R., Rossi-Mori, A., Paton, R., Rector, A., Magliani, F., & de Vries Robbe, P. (1992). The GALEN Dream, *International Classification, 19(4),* 188-191.

Rada, Roy & Murphy, Clare (1992). Searching versus Browsing in Hypertext. *Hypermedia, 4(1),* 1-30.

Rada, Roy (1997a). *Virtual Education Manifesto.* Liverpool, England: Hypermedia Solutions Limited, and available via http://hsl.gnacademy.org/ gnacademy/hsl.

Rada, Roy & Schoening, James. (1997b). Educational Technology Standards. *Communications of the ACM, 40 (9),* 15-18.

Reisman, S., & Carr, W. A. (1991). Perspectives on Multimedia Systems in Education. *IBM Systems Journal, 30(3),* 280-295.

Schank, R. C. (1994). Active Learning through Multimedia. *IEEE MultiMedia,* 1994, 69-78.