

DiLight: a Digital Library based E-Learning Environment for Learning Digital Libraries

Daqing He, Ming Mao, Yefei Peng
School of Information Sciences
University of Pittsburgh
135 North Bellefield Avenue, Pittsburgh, PA 15260
{dah44,mim13, yep3}@pitt.edu

Abstract: Digital library (DL) increasingly becomes an important learning topic for training librarians and information specialists. However, due to many perspectives, models, and activities developed in DL field, it is difficult for students to develop an initial clear understanding of DL. Current available e-learning environments (such as Blackboard) cannot provide the desired functionalities for supporting students' learning on DL. In this paper, based on the integrated functionalities of collecting, organizing, retrieving, and preserving digital objects in a collection, we convert a DL platform into an integrated and interactive e-learning system. The system, called DiLight, concentrates on organizing lecturing slides, videos, reading materials and students comments into meaningful items/documents, and provides multiple retrieval methods that are suitable for students' varied tasks, needs and preferences. Students' initial feedbacks are very positive, and demonstrate the usefulness of DiLight system in their learning processes.

Introduction

A digital library (DL), in general, refers to a collection of digital objects (such as digital text, images, and videos) and a set of associated techniques and services that help to collect, organize, retrieve, and preserve those digital objects for a community of users (Borgman 2000). ACM Digital Library (<http://portal.acm.org/dl.cfm>) is an example of DL where journals, magazines and conference proceedings are organized for people to access remotely. Along with the advancement of information technologies, and various governmental funding initiatives, DL has become a rapidly developing research and practice area. However, on the other hand, since different communities (i.e., computer scientists, library and information scientists, and corporations) participate in the effort and generate many different DL theories, models and projects (Fox & Urs. 2002), the DL field contains many different perspectives of what the digital libraries and related research topics ought to be. Consequently, it makes increasingly difficult for even DL researchers to grasp the whole picture of DL, and develop a clear understanding of the different perspectives, not to mention students who just enter the field (Spink & Cool 1999).

In recent years educating students and librarians on DL has become one of the most important courses taught in many universities (Spink & Cool 1999; Saracevic & Dalbello 2001). However, teaching DL courses faces three problems. First, as stated above, DL is a complex topic. Students entering the field face a steep learning curve when they try to grasp core contents and topics, and understand different DL views and approaches comprehensively. Second, the students are from different backgrounds with different technology preparation. This problem is especially significant in the field of library and information science, where at the one end of spectrum, there are freshly graduated art and humanity undergraduate students who have no library and information technology (IT) training, or experienced veteran librarians coming back for re-training on IT after many years; at the other end of spectrum, there are freshly graduated IT undergraduates, or digital librarians with many years' experience. The third problem is that the instructional delivery is still limited largely to in-class lectures, out-of-class readings, projects, and isolated assignments. The current structure of instructional components leaves much of the intellectual integration to the students.

We believe that an e-learning system can help to easy the first and the third problems. Specifically, a well designed e-learning system should 1) collect fragmented but related course materials, and organize them in such ways that the students' interaction with the materials would be intuitive, meaningful, and effective. 2) manage and present the materials in a dynamic and flexible manner because students have diverse backgrounds and learning

preferences. For example, to complete a learning task, students may request and utilize materials in various forms or access from different angles. 3) allow multiple ways to access materials from any place and at any time, either by browsing or by search. 4) provide community space for students to communicate with and help each other. Overall, the goal of the e-learning system should be able to improve the quality of the students' experience and increase the effectiveness of that experience.

However, conventional e-learning systems (i.e., Blackboard¹) cannot fulfill the above requirements. Though the Blackboard system can hold course related materials and store them in different file folders, it still takes the metaphor of file cabinets, treating materials as unalterable collections with a static structure. It is difficult to be changed by instructors and unchangeable by students. In addition, it lacks of powerful and intuitive search tools and does not support content or recommendation based browsing.

We believe that DL technologies can help in building a better e-learning system. First, a DL contains one or several collection(s) of self-contained digital documents that are accompanied with metadata describing the contents and relevant characteristics of the documents. The collection and the metadata help to group and organize fragmented relevant materials, and at the same time, enable accessing and presenting materials in a dynamic and flexible manner. Second, like a traditional library, a DL is capable to provide a community space for people to communicate and help each other.

In this paper, we talk about the design and development of an integrated, interactive, and effective e-learning system called DiLight. It is based on an open source DL platform called DSpace² to take the advantage of DL's integrated functionalities of collecting, organizing, retrieving, and preserving digital objects in a collection. However, to match the above criteria for a well designed e-learning environment, DiLight has to extend DSpace in many ways. In specific, the design of DiLight system has to consider methods to resolve the following issues:

1. Collection building:
 - a) What should be included in the DiLight's collection? Should the scope of the collection resembles many publication oriented digital libraries, such as ACM Digital Library and Elsevier Digital Library, or a course oriented digital libraries that store DL course related materials, like slides, lecture videos, readings and textbooks?
 - b) What constitutes a document in the collection? For any DL system to perform correctly, documents as the basic logic unit. Related issues also include how to collect and represent those documents.
2. Access Support to faculty and students
 - a) Both faculty members and students would interact with the collection under different tasks, needs, and requirements. Faculty members may want to create, organize, submit, or maintain the course materials, and students may perform actions like locating or writing notes on certain materials. DSpace only supports basic retrieval and browsing functionalities. What are other more advanced access methods that DiLight should support?
3. Community Collaboration
 - a) Students can learn from teachers, or from their fellow students. It is important and feasible for students to build up virtual community in DiLight to help each other's learning. How to help students to collaborate with each other in DiLight, and what tools can DiLight provide?

In remaining of this paper, we will first summarize related work in the literature, then concentrate on presenting our approaches on collection building and access support. We will not cover community collaboration here due to the limitation of space. We will also present some initial responses from an informal evaluation conducted on the group of library and information science students who were learning DL topics. Finally, we will give some conclusion remarks and directions for future work.

Related Work

The research on e-Learning systems has combined research interests and efforts from various fields, and produced diverse systems. Two examples of widely-applied learning management systems are Blackboard

¹ <http://www.blackboard.com>

² <http://www.dspace.org>

(<http://www.balckborad.com/>) and WebCT (<http://www.webct.com/>). There are also e-learning systems developed in academic settings. For example, Minka et al., at MIT Media Lab, developed an interactive program to help students learn how to query visual and image databases (Minka & Picard 1996). Johnson et al., at the University of Southern California, developed animated pedagogical agents to provide face-to-face interaction in learning (Parker & Johnson 1990). Both of them enable students to more actively engage course materials during their learning. Carlson et al., at the University of Colorado Boulder, developed an integrated teaching environment allowing students learn engineering-related courses in one single, well-organized environment (Carlson & Sullivan 1999). Other systems that encourage students' participation in classrooms by providing them an active learning environment can be found in (Parker & Johnson 1990; Meyers & Jones 1993; Paradigm 2001). There are also other global courseware-reusability frameworks providing repositories of reusable educational objects, such as ARIADNE (<http://www.ariadne-eu.org/>) and EdNa (Educational Network Australia <http://www.edna.edu.au/edna/page1.html>).

To support a richer set of educational functions, knowledge representation technologies such as ontology have been integrated into e-learning environments. For example, the Scholarly Ontologies (ScholOnto) project describes an ontology-centered approach to provide an infrastructure for making claims about the significance of research publications (Shum 1998). The virtual hyperbooks model supports collaborative learning by integrating a reusable document repository (called fragments repository) and connects it to a domain ontology developed for the project (Falquet & Ziswiler 2005). Dicheva and Dichev propose a topic-map based system for building, maintaining and using concept-based ontology-aware digital course libraries (Dicheva & Dichev 2004). The Diogene project develops a Web training environment, based on the ACM Computing Classification Scheme (ACM CCS), for classifying books, journal articles and conference proceedings in the field of computing into a four-level subject hierarchy (Vergara et al. 2003). The Courseware Watchdog is an ontology-based tool for finding and organizing learning materials in a decentralized way, in which the ontology serves as the basis for enhancing both the browsing and searching functions inside the system (Tane et al. 2003).

There are several open source digital library platforms, among which Greenstone³, Fedora⁴, and DSpace are most often cited. DSpace was selected as the base for DiLight system because it contains the combination of required modules, and is available at the time for our development. It is a digital repository system developed by the collaboration of MIT Libraries and HP labs. It has the abilities of capturing and describing submitted digital materials, distributing an organization's digital assets over the Web through a search retrieval system⁵. DSpace accepts various formats of digital materials, and models all input materials as items, which consists of a set of bitstreams and their format information. A collection is a group of related items, and a set of collections together is a community.

DiLight E-Learning Environment

DiLight is an e-learning environment for library and information science students to learn DL related topics. As shown in Figure 1, DiLight has extended DSpace in many aspects, including collection building, knowledge representation, access methods, and user annotation and community collaboration. In the remaining of this section, we will concentrate on the extensions at collection building, knowledge representation, and access methods.

DiLight Collection

The goal of DiLight is to provide supports for students to learn DL topics. To best serve students in their learning tasks, the scope of DiLight collection is defined as the course related materials, like slides, lecture videos, readings and textbooks. In this case, we define an item (i.e., a document) in DiLight as a set of topic related course materials such as slides, video and text chunks. Two possible alternatives are 1) one slide or one time segment of videos, or 2) slides or videos for one whole class. None of them are selected here because they either too small to always have self-contained information, or too big that can often correspond to a class lasting for several hours and materials that might not have close relations. We do agree that it is an open question as what constitutes a topic in

³ <http://www.greenstone.org/cgi-bin/library>

⁴ <http://www.fedora.info/>

⁵ More information is available at <http://wiki.dspace.org/EndUserFaq>.

our approach. We decide to let the course instructors to define it because they know the best of the course materials, and students' needs.

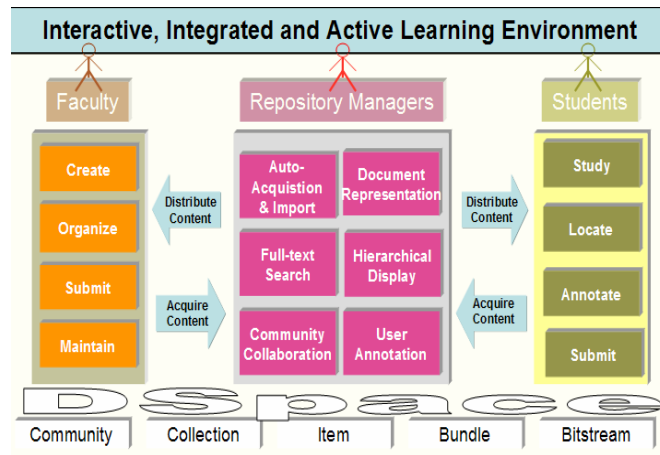


Figure 1 The Architecture of DiLight System

DiLight is capable of handling complex items. An item often contains an HTML page that is the result of converting the content of a set of slides into jpeg pictures. The reason to abandon more straightforward PDF route is because that requires opening extra software to view the slides content. Image version of the slides, though larger in size, preserves the graphical presentation of the slides and is able to be displayed in all modern browsers without extra downloads. With the help of Carnegie Mellon University's CourseCast system, most items are accompanied with the in-class delivering recorded on digital video, where the videos and their corresponding slides are associated with each other based on slide turning information. Each item has its associated metadata, which consists of the standard Dublin Core element set.

Based on the consideration of the reality of building a collection of course materials, DiLight supports two modes of constructing collections. The collection can be incrementally expanded by adding individual items, or through a batch mode of importing large number of items at one time. The incremental expansion is done by utilizing the existing DSpace graphical interface, where users can specify the related stream files and filling in corresponding metadata. Our work here is on batch mode. We have developed mechanism that can read from an XML file the specifications of a set of items and corresponding metadata, and automatically locate the files and uploaded into DiLight collection.

DiLight's DL Course Ontology

In e-learning domain, though standard schemas like LOM (Learning Object Metadata)⁶ and Dublin Core⁷ have been applied to describe learning objects, these standards only focus on the minimal set of attributes (i.e., type of object, author, owner, terms of distribution, and format). Such a simple structure cannot help students' learning complex knowledge and relationships among topics. Therefore, we have explored ontology as a more comprehensive approach to integrate the content, structure and evolution of the learning materials. An ontology, as "a formal explicit specification of a shared conceptualization" (Gruber 1993), explicitly represent concepts and their relationships in a logical and machine-interpretable form, and thus enable automated inference capability over the concepts. Current research on ontologies has shown that they do facilitate the retrieval, interaction and management of resources (Horrocks & Hendler 2002; Studer & Staab 2003).

After examining several well known ontology like ACM Computer Classification System (ACM CCS)⁸ and SWEBOK ontology⁹, we found that they are not suitable for our situation because their coverage is too generic.

⁶ <http://ieeeltsc.org/wg12LOM/>

⁷ <http://dublincore.org/>

⁸ <http://www.acm.org/class/1998/>

⁹ <http://www.swebok.org/>

For example, “Digital Libraries” is at third-level under “Information Systems” and “Information Storage and Retrieval” in ACM CCS without any further detailed classification. Therefore, we took the route to build our own course-oriented DL ontology. There is no “correct” way or methodology to model a domain into an ontology (Noy & McGuinness 2001). Different purposes and tasks usually end up with different ontologies. However, we do identify that our ontology should satisfy three requirements: 1) to organize and visualize major topics of the DL course through hierarchical and association relationship, 2) to support students’ access to materials relevant to certain topics, and 3) to allow future integration with other courses on DL or related domains.

We observe that, although actual contents of courses vary substantially, there are always common structures among different courses and the relationships between these structures are relatively stable. Therefore, we design a relatively small set of conceptual and relational types to capture the essential structure existing in courses. Our goal is to make the types and structures simple to understand, yet expressive and extensible enough to multiple domains so that the ontology can be broadly adopted and easily applied later. For instance, we notice that a course usually consists of several “Discussion Points”, which can be related to “Topics”, “Standards”, “Applications”, “Technologies and Techniques”, and so on. Figure 2 illustrates main classes and relations created in our DL ontology, which currently contains 16 classes (i.e., applications, discussion points, issue, model and framework, theory and methodology, topic, standards, statements, etc.), 43 relations (i.e., apply, based_on, derived_from, discuss, has_issues, illustrate, modify, propose, extend, etc.).

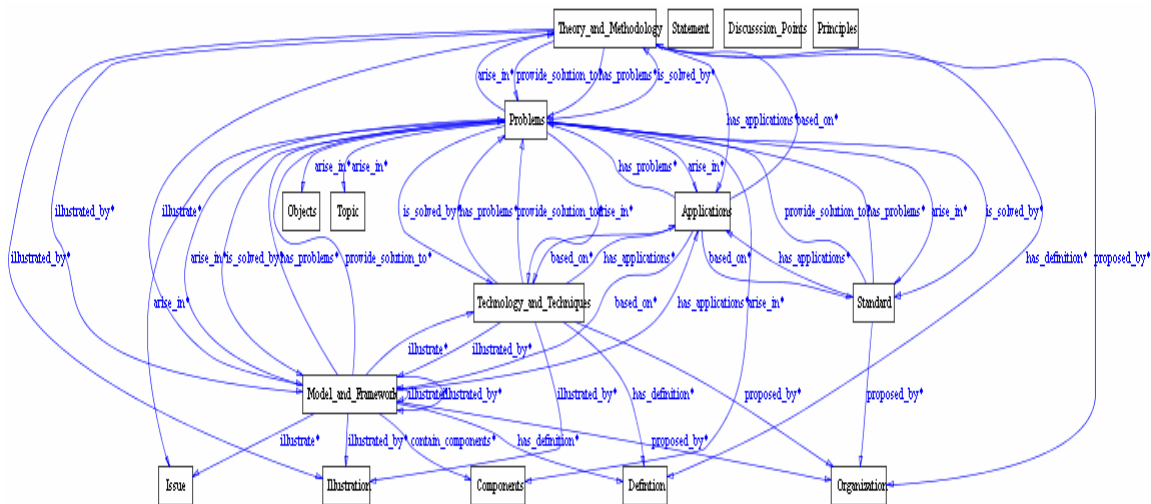


Figure 2 Main classes and relations in the DL ontology, generated by Protégé 3.1

DiLight’s Multiple Access Methods

Users’ information access methods can be categorized as either passively receiving updates or active seeking information. Recommendation and filtering are two commonly used technologies for the former, and search and browsing are for the latter. DSpace framework provides some basic access methods, including email signup for receiving collection updates, and simple search methods by title, author, identifier, and full content. Users can also browse the collections by title, author, and dates.

DiLight extension over DSpace’s access methods has been concentrated on four areas: 1) lecture-based browsing; 2) ontology-based browsing, 3) ontology-based search, and 4) association recommendation.

Lecture-based browsing. A course contains a serial of lectures. Students are often required to access materials by lectures, or remember course topics by lectures. Therefore, some form of lectures and the clear association between the lectures and course materials should be presented in DiLight system. We utilize the “collection” level of document groups in DSpace to represent each lecture. Users can browse through the lecture collection hierarchy to locate certain topics within a particular lecture.

Ontology-based browsing. Dicheva and Dichev (Dicheva & Dichev 2004) stated that students are often unaware of the complete context of a learning task, and need help in getting oriented in the conceptual structure of the subject domain. The ontology we constructed does provide a visual presentation of such conceptual structure about DL course topics. With the 43 types of relationships provided in the ontology, students can navigate through the hierarchical structure, exploring or discovering internal semantic relationships among course topics/concepts. For example, from topic “XML” it is possible and reasonable for students to jump to topic “Semantic Web.”

Ontology-based search. In an e-Learning environment, one of the biggest problems that students face is to use what terms or keywords when searching course materials. Simple keyword queries are valuable when students know exactly what they are seeking for and/or the information is well-defined. However, simple keyword searches does have the limitation of returning false relevant documents containing homonyms (i.e., “organization” used in collection building and in communities), and missing truly relevant documents that have only synonyms (i.e., “architecture” vs. “framework”) or abbreviations (i.e., “Digital Library” and “DL”). Because teachers and students may have different viewpoints and knowledge levels of learning materials, the term mismatch problems are common in learning environment. In this case, an ontology-based search can overcome the problem by utilizing its built-in inference engine to perform deduction or other forms of automated reasoning to locate documents that are semantically similar to the query terms regardless of the surface difference, thus achieve more accurate or comprehensive results. For instance, our collection contains three documents concerning data encryption. Document A is about MD5 (Message-Digest Algorithm 5), B is about RSA (a public-key encryption technology), and C is about DES (Data Encryption Standard). A is assigned to node “Algorithm in Data Transmission”. B and C are assigned to node “Technique and Standard in Data Transmission”, which is related to “Algorithm in Data Transmission”. A normal keyword search on “RSA” would only return B. However, our ontology-based search can also return A and C with the help of the inference that the nodes of A and C are related to that of B, even though neither A nor C contains keyword “RSA”. Another advantage of ontology-based search is that it provides facilities to explicitly specify the scope of a search. Scope-based search is useful especially when the user has some knowledge about areas/scopes in which the information s/he is looking for might be. A clear defined scope helps to reduce ambiguities in query terms and thus increase search accuracy. For example, if the student knows that her/his interest in “RSA” is about security issues in DL, not related to other topics, s/he can explicitly identify the search scope as within topic “Security Issues in DL” instead of searching the whole data repository of DiLight.

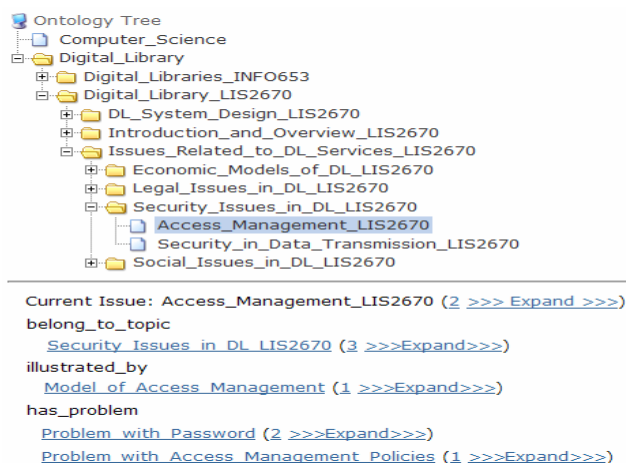


Figure 3 Ontology-based Information Access in DiLight

Associated recommendation. Content-based browsing enables access to materials via links between parent and child nodes, whereas ontology-based search provides a set of access points over the whole ontology. However, both of them are limited at providing navigation beyond the immediately linked topics. Sometimes such connection between topics over long distance is important for students to build up understanding about related topics that are taught in different classes or under different circumstances. Especially where their understanding of a topic is low, they need to be directed intelligently towards resources of relevance. For example, it is important for students to understand that the discussion of different communities in DL under the lecture “Introduction of DL” is related to the different economic models for DL under the lecture “Economic Issues of DL”. In this situation, with a

comprehensive understanding of internal connections between different topics, the teacher can simply utilize the “derived from” relationship in our ontology to explicitly markup the connection. Thus, an associated recommendation is provided when students access either of the topics. With such relationship, documents in both topics are linked together across the ontology.

Initial Responses from Students

A group of 11 Library and Information Science students were asked to explore the usability and functionality of the DiLight system under the understanding that the system will evolve to be a learning support system for their DL studies. The questions that they should keep in their minds during the exploration include those about collection building (e.g., which information available in the system is useful, might be useful, or not useful at all? Which useful information is missing?), about methods for accessing materials (e.g., are the browsing and search functions adequate for your study?), about material organization and representation (does the arrangement of the content make sense to the needs of your study?), and about interface and interaction design (e.g., does the layout of the interface make sense to the needs of your study?).

Students in general agree that DiLight system contains most useful materials for their studies. As one student states the system “centrally locates information from many sources which are useful to me: articles, videos, presentations, website examples, etc. It is comprehensive on each topic.” 100% of students agree that the browse and search capabilities provided in the system are adequate in supporting their learning tasks. Additionally, 73% (8/11) of students state that ontology presentation provides a great organizational tool to find information. They like the ontology because it simulates users’ presumed knowledge of the “desktop metaphor” and create “familiarity” to their topics. With the help of the ontology, their learning curve for interacting with the materials is short and painless. However, we do have one student felt confused about ontology because he/she did not understand what the word “ontology” means. Two other students stated that they need a bit more time to learn how to use ontology to access information, therefore they preferred using search function. This different response demonstrates the usefulness of having multiple access methods in DiLight. Finally, 100% of students think the layout of the interface is easy to learn and the arrangement of the content is logical and self-explanatory. One consequence is, as one student stated, that the layout “makes it relative easy to remember” the content. Some trivial problems of DiLight mentioned by students include how to rank and display the search results and adding more related materials into the repository. Overall, among the students (3 out of 11) who made direct comparison between DiLight and current CourseWeb/Blackboard system, all of them think that DiLight is a great improvement.

Conclusion and Future Work

In this paper, we have discussed the design and development of DiLight system. Based on DSpace DL platform, the DiLight system has been extended into an integrated and interactive e-learning environment. Our discussions concentrate on collection building, knowledge representation, and access methods in the DiLight system. To better serve students in their studies, slides and videos that are related to one course topic are grouped together to be an item/document inside DiLight’s collection. The exact definition of a topic in a course is left to the instructor of the course to define. Course related concepts and topics are also arranged into hierarchical structure with associated relationship, which in the whole is a DL course ontology. The structure of the ontology is kept generic for easy expanding to other course domains. Multiple access methods, include lecture-based browsing, ontology-based browsing and searching, and associated recommendations, are integrated into DiLight to enable users with different background and preference to have an appropriate access method. The preliminary survey from students shows that: DiLight system contain most required materials for students’ studies on DL and related courses. The lecture and ontology based browsing and retrieval methods are useful for students to access materials. The whole package provided by DiLight system makes it a great improvement over blackboard system or many other commonly used course support systems

There are several possible directions for future work. We are expanding DiLight collection with more course materials from other DL courses or related courses. Our aim is to make DiLight collection richer in content and more comprehensive in domain coverage. We are also revising the DL course ontology by incorporating more DL concepts and their relationships in. With the integration of other related non-DL courses, we will examine the ontology for covering several courses. We are also interested in further enhancing the ontology-based information

access methods. Last but not least, we are working on building social navigation and online commenting facilities to allow students to collaborate with each other in their learning process, which makes DiLight system an open and collaborative e-learning environment.

Acknowledgement

This work is supported by the Provost's Innovation of Teaching Award at University of Pittsburgh.

Reference

- Borgman, C., 2000. *From Gutengerg to the Global Information Infrastructure*: MIT Press.
- Carlson, L. e. & J. F. Sullivan, 1999. Hands-on Engineering: Learning by Doing in the Integrated Teaching and Learning Program. *International Journal on Engineering Education*, 15, 20-31.
- Dicheva, D. & C. Dichev, 2004. A Framework for Concept-based Digital Course Libraries. *Journal of Interactive Learning Research*, 15(4), 347-64.
- Falquet, G. & J.-C. Ziswiler, 2005. A Virtual Hyperbooks Model to Support Collaborative Learning. *International Journal on E-Learning*, 4(1), 39-56.
- Fox, E. A. & S. R. Urs., (2002). Digital Libraries, in *Annual Review of Information Science and Technology*, 503-89.
- Gruber, T., 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2), 199-220.
- Horrocks, I. & J. Hendler, (2002). The Semantic Web - ISWC 2002, in *Proceedings of the First International Semantic Web Conference*, eds. I. Horrocks & J. Hendler Sardinia, Italy: Springer Verlag.
- Meyers, C. & T. B. Jones, 1993. *Promoting Active learning: Strategies for the college classroom*: Jossey-Bass Publishers.
- Minka, T. P. & R. W. Picard, (1996). Interactive Learning with a 'society of models', in *Conference on Computer Vision and Pattern Recognition*.
- Noy, N. & D. McGuinness, (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*, Stanford Knowledge Systems Laboratory.
- Paradigm, T., 2001. Course Participation: An Active Learning Approach Employing Student Documentation. *Journal of Marketing Education*, 23, 187-94.
- Parker, L. M. P. & R. E. Johnson, 1990. Does Order of Presentation Affect Users' Judgement of Documents ? *JASIS*, 41(7), 493-4.
- Saracevic, T. & M. Dalbello, (2001). A survey of digital library education, in *American Society for Information Science and Technology, (ASIST 2001)*, 209-23.
- Shum, S. B., 1998. Evolving the Web for Scientific Knowledge: "First Steps Towards an "HCI Knowledge Web". *Interfaces Magazine*.
- Spink, A. & C. Cool, (1999). Education for Digital Libraries, in *D-Lib Magazine*.
- Studer, R. & S. Staab, 2003. *Handbook on Ontologies in Information Systems*: Springer Verlag.
- Tane, J., C. Schmitz, G. Stumme, S. Staab & R. Studer, (2003). The Courseware Watchdog: an Ontology-based tool for finding and organizing learning material, in *Fachtagung Mobiles Lernen und Forschen*.
- Vergara, M., N. Capuano & E. Sangineto, (2003). Diogene: a Training Web Broker for ICT Professionals, in *2nd Int. Conference on Multimedia and ICT in Education (m-ICTE 2003)* Badajoz, Spain, 330-4.