The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1468-4527.htm

Refereed article received
19 June 2009
Approved for publication
10 March 2010

Supporting information access in e-learning by integrating digital libraries and ontology

Daqing He and Yefei Peng
School of Information Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
Ming Mao
SAP Labs, Palo Alto, California, USA, and
Dan Wu
School of Information Management, Wuhan University, Wuhan, People's Republic of China

Abstract

Purpose – The purpose of this paper is to examine the integration of digital library (DL) technologies with ontology-based knowledge representation in providing semantic rich information access (IA) in e-learning. DL technologies have powerful and flexible content management and access functionalities, whereas ontology helps teachers and students to link content materials to their learning objectives. This paper demonstrates that the integration provides a powerful and meaningful e-learning environment.

Design/methodology/approach – DiLight is designed as an interactive e-learning system that integrates DL and ontology technologies. By conducting comparative experiments involving DiLight in students' actual learning process, the authors examined the advantages and limitations of DiLight in e-learning.

Findings – Compared to a widely used e-learning environment, DiLight can provide significantly better support for students' complex IA tasks because DiLight is more useful for relationship discovery and problem solving. DiLight is also effective even when students were either less familiar with tasks or felt that they were more difficult. There is no single best access method for all learning situations. Therefore, multiple IA methods should be built into e-learning systems. Although most of time the search was the first choice of the students, ontology-based methods were useful in supporting them to complete their tasks too.

Originality/value – This is a comparative empirical study using an interactive e-learning system called DiLight to explore the usage of integrated DL and ontology in e-learning. The experiment results demonstrate the value of the multiple IA methods provided by DL, and the usefulness of integrating DL with ontology.

Keywords E-learning, Digital libraries, Information, Facilities

Paper type Research paper

Online Information Review
Vol. 34 No. 5, 2010
pp. 704-728

1468-4527
DOI 10.1108/14684521011084582

This work is partially supported by the Provost's Innovation of Teaching Award at University of Pittsburgh and by the National Social Science Foundation of China under the agreement 09CTQ026.

Introduction

With the advancement of computer, multimedia and network technologies, alternatives to traditional classroom learning have been developed. E-learning is one such alternative where students can access course-related materials via online computer systems. It is being explored as an effective way of delivering materials to previously unreachable students with previously unavailable access and presentation methods. Horton (2000, p. 6) even claims that it is "part of the biggest change in the way our species conducts training since the invention of the chalkboard or perhaps the alphabet".

However, e-learning systems still face several challenges in terms of content organisation, knowledge representation (KR), and access management. First, the document collections in e-learning systems contain complex and often mixed-media materials because of the widespread application of digital and presentation technologies in modern learning activities. At the same time, the collections are constantly updated with materials being added, edited, or removed, often by teachers, but sometimes by students too if collaborative activities from students are encouraged. Therefore, the collection organisation side of e-learning systems should be flexible in content representation, friendly in content management, and intuitive in user interface. Second, along with the advantage of allowing students to access the learning systems at any time and in any place, the presence of teachers during the learning process, which is an important component of face-to-face education, could be missing. How to retain the knowledge provided by teachers in an e-learning environment and present it in a straightforward and intuitive way is an important KR challenge that e-learning systems need to address. Third, although the ways of accessing and interacting with learning materials have traditionally been dictated by teachers, students in e-learning systems can choose their own strategies and means to do so. This therefore poses access management challenges in that it requires e-learning systems to have multiple access and interaction methods to accommodate students' various needs and preferences. Of course, we acknowledge that e-learning systems are facing many other challenges too. For example, Conlan *et al.* (2006) talk about the challenges of bridging the gap between engineering sound e-learning systems with pedagogically appropriate and effective theories, and Brusilovsky and Peylo (2003) present studies aimed at the challenges of adding adaptation and intelligence into e-learning systems.

The research presented in this paper aims to address the above three challenges by integrating into e-learning systems the technologies from one related discipline: digital libraries (DLs). A 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, organise, retrieve, and preserve those digital objects for a community of users (Borgman, 2000). Documents in the collection are self-contained digital materials accompanied by metadata describing the content and characteristics of the documents. In DL, the roles of the collection and the metadata are to help to group and organise fragmented-related materials into meaningful and logical units, and at the same time, enable access to and presentation of those materials in a dynamic and flexible manner. Therefore, DL technologies may provide an answer to the content representation challenge mentioned above. Besides combining DL technologies, our approach also explores the application of KR technologies – specifically ontology – and information retrieval techniques in resolving the content organisation and access challenges.

In summary, the goal of this research is to study the integration of DL technologies, with the help of ontology and information retrieval technologies, into e-learning systems to improve the quality of the students' learning experience and to increase the effectiveness of that learning process. Specifically, our work concentrates on exploring answers to the following three research questions:

RQ1. With the help of DL technologies, how can fragmented, mixed-media based learning materials in e-learning systems be organised in such way that the students can interact with the materials in an intuitive, meaningful and effective manner?

RQ2. With the help of DL and KR technologies, how should learning materials be represented and organised to retain teachers' knowledge about course topics and materials?

RQ3. With the help of DL, ontology-based knowledge representation, and information access (IA) technologies, what are the multiple access methods that enable students with diverse backgrounds and learning preferences to access the collection using their own effective methods?

Our research method is practical and experimental. With the help of technologies from DL, KR, and IA, we first developed an e-learning system using an open source DL system as the foundation. We then expanded the system with enhanced content organisation, knowledge representation and information retrieval capabilities. The final system, which is called the DiLight system, was then tested in experimental settings to examine whether our approach is valid or not.

In the remainder of this paper we will present our approach in detail. The paper begins with a review of previous related work in the field of e-learning, then presents important features of DiLight, paying specific attention to its collection building, ontological representation of the DL course topics, and the multiple IA methods developed for supporting various students' needs and preferences. We then talk about the experiments we conducted to evaluate the usefulness and usability of DiLight in supporting students' learning. The discussion of the results is followed by a conclusion about the importance of the work and some potential future directions.

Related work

Many people believe that new and emerging technologies will generate great advances and improvements in learning and instruction (Spector, 2001). This is why e-learning has received considerable attention from both academic researchers and practitioners. Some researchers called the recent wave of studying e-learning an e-learning revolution (Galagan, 2000). Specifically, e-learning has been seen as the technology to provide advantages such as consistent and worldwide training, reduced delivery cycle time, increased learner convenience, and reduced information overload (Welsh *et al.*, 2003).

E-learning has become an increasingly popular mode of instruction in higher education due to the continual advances in internet and multimedia technologies. Although early definitions of e-learning included electronic devices such as radio broadcasting and teleconferencing (Siritongthaworn *et al.*, 2006), researchers have increasingly redefined the meaning of e-learning as "the use of internet technologies to deliver a broad series of solutions that increase knowledge and performance" (Rosenberg, 2001, pp. 28-9).

Because of the nature of e-learning, its literature can be roughly classified into the theories behind e-learning and practices of e-learning.

Most e-learning theories borrowed ideas from the related concept of distance learning (Gunasekaran *et al.*, 2002). Researchers found that the principal advantages of distance learning and e-learning are "flexibility, cost savings, and more efficient use of time and staff motivation" (Gunasekaran *et al.*, 2002, p. 47). As Gunasekaran *et al.* (2002) point out, people in e-learning environment can master the learning materials more effectively because it is possible for them to learn at their own pace and therefore retain more information and become more proactive. Henry (2001, p. 249) examined the factors that affect the successful implementation of e-learning, and stated that "e-learning needs to be compelling to the audience it targets, offering the learner a resource that is seen to be appealing, valuable and productive to their goals and aspirations". From the opposite angle, Alexander and McKenzie (1998) discussed the reasons that e-learning could fail, which include failing to set clear goals/objectives and inappropriate training for students.

Developing e-learning systems has combined research interests and practical efforts from various fields, and produced diverse systems. We have seen many widely used commercial learning support and management systems, of which Blackboard and WebCT (www.webct.com/) are two salient examples. E-learning is also a powerful tool for delivering many and varied instructional technologies and methods (DeRouin *et al.*, 2005). For example, Conlan *et al.* (2006) examined the relationship between the Knowledge space theory and the Adaptive Personalised eLearning Service, and Mwanza and Engestrom (2003) presented their experience of using activity theory in the Lab@Future project. Parker and Johnson (1990) at the University of Southern California developed animated pedagogical agents to provide face-to-face interaction in learning. Minka and Picard (1996) at MIT Media Lab developed an interactive program to help students learn how to query a visual and image database. Carlson and Sullivan (1999) at the University of Colorado Boulder developed an integrated teaching environment allowing students to learn engineering-related courses in a single well-organised environment. Several studies discuss systems that encourage students' participation in the classroom by providing an active learning environment (Meyers and Jones, 1993; Paradigm, 2001; Parker and Johnson, 1990).

An important development in e-learning systems was the introduction of adaptive interactions. Brusilovsky (1999) discussed the goal of moving e-learning systems beyond a network of static hypertext pages and developing advanced web-based educational applications that can offer adaptivity and intelligence. Some of the systems focus on the immediate adaptation of instructions to provide timely and targeted feedback for the students and the instructor, e.g. using a digital dashboard (Brown *et al.*, 2006) and the Adaptive Personalised eLearning Service (Conlan *et al.*, 2006), whereas some other systems pay attention to visualisation, e.g. Ahn *et al.* (2006) and social navigation, e.g. Mertens *et al.* (2006) and Nokelainen *et al.* (2002).

Although published too long ago to include recent e-learning work, there exist several well-received literature reviews of e-learning practice and research (Burgess and Russell, 2003; Kosarzycki *et al.*, 2003; Salas *et al.*, 2002; Welsh *et al.*, 2003).

Interested readers should consult these papers for more in-depth discussions of e-learning theories and practices.

Because of the widespread application of digital and presentation technologies in learning activities, the collections in e-learning systems often contain complex and mixed-media materials. This motivated people to look for more powerful content management tools. The MIxed media Networked Digital library (MIND) project is one example of such activities aiming to explore efficient and effective mechanisms for indexing and retrieving mixed media information from a DL of theses (Acosta-Diaz *et al.*, 2006). Their focus is on indexing mixed-media documents, so the study of supporting learning activities is relatively peripheral. DL technologies have been used to support other types of learning activities too. The National Science Digital Library is one important example of such applications of DL in learning support (Zia, 2001). It is not an e-learning system, rather it is a digital repository of collections of learning materials, which can be easily accessed and adopted for various science, technology, and mathematical learning activities.

To support a richer set of educational functions, knowledge representation technologies such as ontology have been integrated into the e-learning environment. For example, the Scholarly Ontologies (ScholOnto) project describes an ontology-centred 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 connecting it to a domain ontology developed for the project (Falquet and Ziswiler, 2005). Dicheva and Dichev (2004) propose a topic-map based system for building, maintaining, and using concept-based ontology-aware digital course libraries. The Diogene project developed a web-training environment, based on the ACM Computing Classification Scheme, for classifying books, journal articles, and conference proceedings in the field of computing into a four-level student hierarchy (Vergara *et al.*, 2003). The Courseware Watchdog is an ontology-based tool for finding and organising learning materials in a decentralised way, in which the ontology serves as the basis for enhancing both the browsing and searching functions inside the system (Tane *et al.*, 2003).

Our research is unique, when compared to the above studies, in that it concentrates on expanding the capabilities of e-learning systems in three important aspects:

(1) It expands the content representation aspect by integrating DL technologies.

(2) It expands the content organisation aspect by integrating DL and ontology based KR technologies.

(3) It expands the content access aspect by integrating ontology and IR technologies.

These three integrations help to build a solid foundation for supporting learning activities. By implementing these three integrations, the DiLight system has powerful content management and access capabilities.

DiLight system

As mentioned already, our research interests focus on flexible content management, rich knowledge representation, and powerful IA capabilities in e-learning systems. DiLight was designed to reflect these emphases. With the help of an open source DL platform called DSpace (www.dspace.org/), DiLight takes advantage of the DL's integrated functionalities of collecting, organising, retrieving, and preserving digital objects in a collection to provide an integrated, interactive, and effective e-learning environment. The reason that DiLight uses DSpace rather than several other open source DL platforms, such as Greenstone (www.greenstone.org/cgi-bin/library) and Fedora (www.fedora.info/) is because DSpace contains the combination of required modules, and was available at the time of our development. DSpace is a digital repository system developed through the collaboration of MIT Libraries and HP Labs. It has the abilities of capturing and describing submitted digital materials, and distributing an organisation's digital assets over the web through a retrieval system. It accepts various formats of digital materials, and models all input materials as items which consist of a set of bitstreams and their format information. A collection is a group of related items, and a community in DSpace is a set of collections.

Although DiLight is designed to be a generic e-learning support system, we need a specific learning domain to demonstrate and examine its characteristics. We chose a DL course for library and information science (LIS) students at the University of Pittsburgh as the domain. Besides the obvious reason that one of the authors has been teaching that course, this selection was motivated by the fact that courses educating information specialists including librarians on DL are among the most important topics taught in many universities in recent years (Saracevic and Dalbello, 2001; Spink and Cool, 1999).

Figure 1 shows the conceptual architecture of the DiLight system. The various modules and knowledge bases are there to extend DSpace to resolve issues such as collection building, knowledge representation, access methods, user annotation,

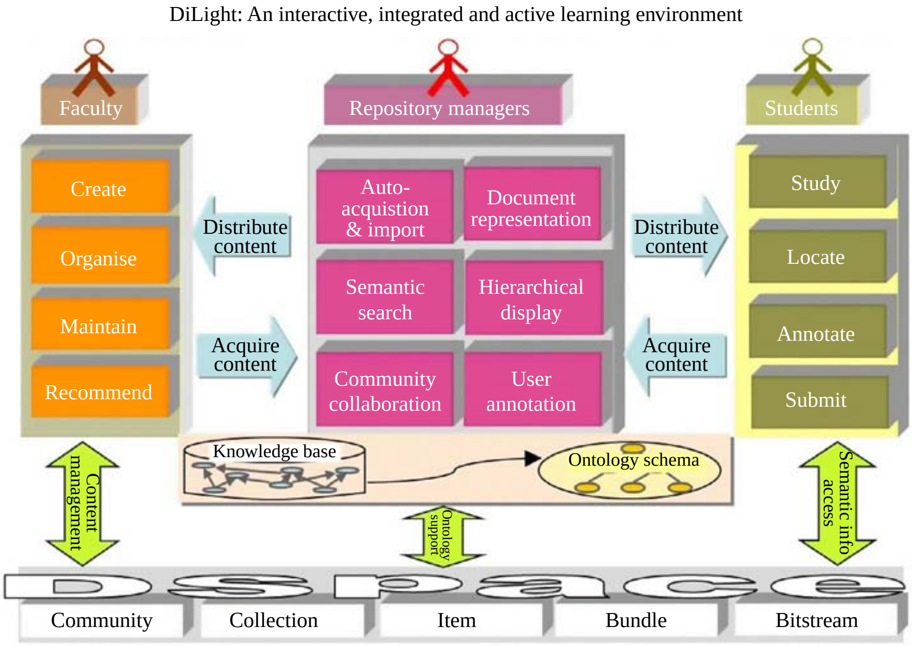


Figure 1.
The conceptual architecture of DiLight system

and community collaboration. In the remainder of this section, we will concentrate on the extensions to collection building, knowledge representation, and access methods, because they are essential to the understanding of the content management and IA capabilities in DiLight.

Content representation in DiLight

DiLight is an e-learning system, but at the same time, it is also a DL system. In a DL system, it is important to define what a document is and the scope of the collection of documents. To best serve students in their learning tasks, the scope of the DiLight collection is defined as the course-related materials, such as 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, videos and text chunks – that talk about the same topic in the course. Potentially, there are two other alternatives: one of which selects one slide or one time segment of videos as a document, and the other treats all slides or videos for one whole lecture as a document. However, we decided to rule them out because they are either too small to always have self-contained information, or too big and too long to contain closely related materials. We do acknowledge that our choice raises the question of what constitutes a topic. To cope with this problem, we decided to let teachers define it as they are most familiar with the course materials and the students' needs.

With the help of DSpace, DiLight is capable of handling the documents in DiLight which are complex items. A DiLight document contains an HTML page that is the result of converting the content of a set of slides into jpeg pictures (Figure 2). The reason to ignore the more straightforward PDF route is that that approach requires opening extra software to view the document content, whereas the image version of the documents, though larger in size, is able to be displayed in all modern browsers without extra downloads and at the same time preserves the graphical presentation. Besides the images of slides, the HTML page also contains all text presented in the original slides, which makes the page searchable.

With the help of Carnegie Mellon University's CourseCast software, most slides in the DiLight collection are accompanied with an in-class delivery of the slide content recorded in digital video format. The software synchronises the slides and their corresponding videos based on slide turning information. The video provides extra

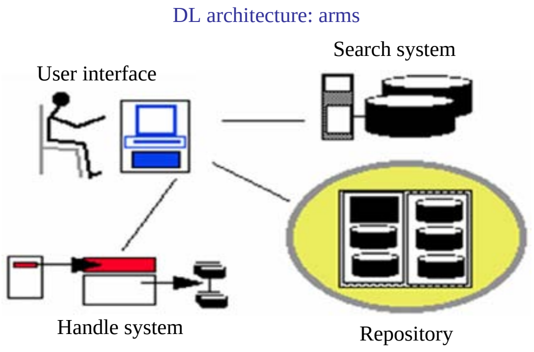


Figure 2.
An example of a slide transferred into a document

details of the document that could be missing in the slides; therefore, one of the future tasks is to add automatic transcriptions of the video content into the document representation.

Each document has its associated metadata, which consists of the standard Dublin Core element set. The elements utilised in DiLight are title, author, issue date, series, URI, and abstract (Figure 3).

With the help of DL features, the construction of the DiLight collection does not have to pay much attention to issues such as how the documents should be structured because it will be handled by the access methods provided in DiLight. Therefore, the remaining issue about the DiLight collections is how documents can be uploaded.

Based on consideration of the reality of building a collection of course materials, DiLight supports two modes of uploading documents. Individual documents can be added through an existing DSpace module, which contains an interface for users to specify the related materials to a given document and fill in the corresponding metadata. A large number of documents can also be uploaded through a batch process designed as a special feature of DiLight. It invites the teacher to participate in the collection building process by writing a configuration file that indicates the start and the end slide numbers of a document, the URL of the corresponding video if available, and the related metadata (Figure 4). The configuration file is in XML format so that a software tool we developed can automatically locate the related files and upload them into the DiLight collection.

Ontological representation of course knowledge in DiLight

Although standard schemas such as Learning Object Metadata and Dublin Core have been used to describe objects in e-learning, and the latter exists in the DiLight system, these standards only focus on a minimal set of attributes (i.e. type of object, author, owner, terms of distribution, and format) which cannot enhance students' learning of complex knowledge and relationships among topics. Therefore, we explored using ontology as a more comprehensive approach to integrate the content, structure, and relationships of the learning materials. Integrating knowledge representation techniques in e-learning has been explored by previous studies (Henze *et al.*, 2004);

Title: Arms DL Architecuture
Authors: Daqing He
Issue Date: 6-Jan-2006
Series/Report no.: Digital Library
Week2 - Architecture of Digital Library
Abstract: DL Architecture: Arms Components of a DL: Arms Repositories Collections of digital objects A DL may have many repositories Each digital object has a handle Metadata The object itself Handle system given the handle for a digital object, returns the info of where it is in the repositories Handle General purpose Identifiers for Online Digital Objects Is handle just another way of saying "unique file name"? Again, why are semantic names so disruptive to the handle system? Criteria for such identifiers Unique, Persistent, Universal Easy to obtain and resolve Real life identifier Social Security Number, National Identification Number driver license number A Question to Think Can URL act as the identifiers of online digital objects? Components of a DL: Arms (continued) Search system Help users to access the repositories May support a variety of protocols User Interfaces End user access the DL System admin manage the DL
URI: http://hdl.handle.net/2199/136
Appears in Collections: Week2 - Architecture of Digital Library

Files in This Item:

File	Description	Size	Format	
contents.html		0Kb	HTML	View/Open

Figure 3.
The metadata of a document in DiLight

```
  <title>Unit 02 - Architecture of Digital Library</title>
  <jpgdir>C:\dspace_files\jpg\2670-u2-arch</jpgdir>
  <rtffile>C:\dspace_files\rtf\2670-u2-arch.rtf</rtffile>
- <document>
    <title>Agenda and Goals</title>
    <slides>1-4</slides>
  </document>
- <document>
    <title>Griffin Three Dimensions of DL</title>
  - <dublin_core>
      <dcvalue element="relation"
        qualifier="belong_to_topic">Griffin_Three_Dimensions_of_DL</dcvalue>
    </dublin_core>
    <slides>5</slides>
  </document>
- <document>
    <title>Definition of DL Architecture</title>
  - <dublin_core>
      <dcvalue element="relation"
        qualifier="belong_to_topic">Definition_of_DL_Architecture</dcvalue>
    </dublin_core>
    <slides>17-19</slides>
  </document>
- <document>
    <title>Arms DL Architecture Model</title>
  - <dublin_core>
      <dcvalue element="relation"
        qualifier="belong_to_topic">Arms_DL_Architecture_Model</dcvalue>
    </dublin_core>
    <slides>20-24</slides>
  </document>
```

Figure 4.
A configuration file for the DiLight collection

however, our approach involves a specific ontology rather than a borrowed generic ontology.

An ontology, as "a formal explicit specification of a shared conceptualization" (Gruber, 1993, p. 199), explicitly represents concepts and their relationships in a logical and machine-interpretable form, and thus enables automated inference capability over the concepts. Current research on ontology has shown that it facilitates the retrieval, interaction, and management of resources (Horrocks and Hendler, 2002; Studer and Stabb, 2003). After examining several well-known ontologies such as ACM Computer Classification System and SWEBOK ontology, we found that they are not suitable for our situation because their coverage is too generic. For example, "Digital Libraries", which is the top concept in DiLight, is at the third level under "Information Systems" and "Information Storage and Retrieval" in the ACM Computer Classification System without any further detailed classification. Therefore, we built our ontology and decided that it should be a DL course-oriented ontology, which aims at describing the DL concepts and their relationships discussed in the DL course.

There is no "correct" way or methodology to model a domain into an ontology (Noy and McGuinness, 2001). Different purposes and tasks usually end up with different versions of ontology. However, we identified that our ontology should satisfy three requirements:

1. organise and visualise major topics of the DL course through hierarchical and association relationships;
2. support students' access to materials relevant to certain topics; and
3. allow future integration with other courses on DL or related domains.

Although the actual content of courses varies substantially, we observe that there is always a set of common concept classes shared among different courses, and the relationships among the concept classes are often relatively stable. For instance, a course usually consists of several "discussion points", which can be related to "topics", "standards", "applications", "technologies and techniques", and so on. Therefore, we designed 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 enough to be understood, yet with adequate expressive and extensible power so that the ontology can be broadly adopted and easily applied later in other course domains. Figure 5 shows some major classes and relations created in our DL ontology, which currently contains 16 classes (i.e. applications, discussion points, issues, models and frameworks, theories and methodologies, topics, standards, statements, etc.) and 43 relations (i.e. apply, based_on, derived_from, discuss, has_issues, illustrate, modify, propose, extend, etc.).

Multiple access methods in DiLight

Users' IA methods can be categorised as either passively receiving updates or actively seeking information. Recommendation and filtering are two commonly used technologies for the former, as search and browsing are for the latter. The DSpace framework provides some basic passive methods, including e-mail signup for receiving collection updates, and simple keyword search by title, author, or full content. Users can also browse the collections by title, author, or dates.

With supporting students' various access requirements in mind, DiLight's extensions to access methods have been concentrated on four areas:

1. lecture-based browsing;
2. ontology-based browsing;
3. ontology-based search; and
4. association recommendation.

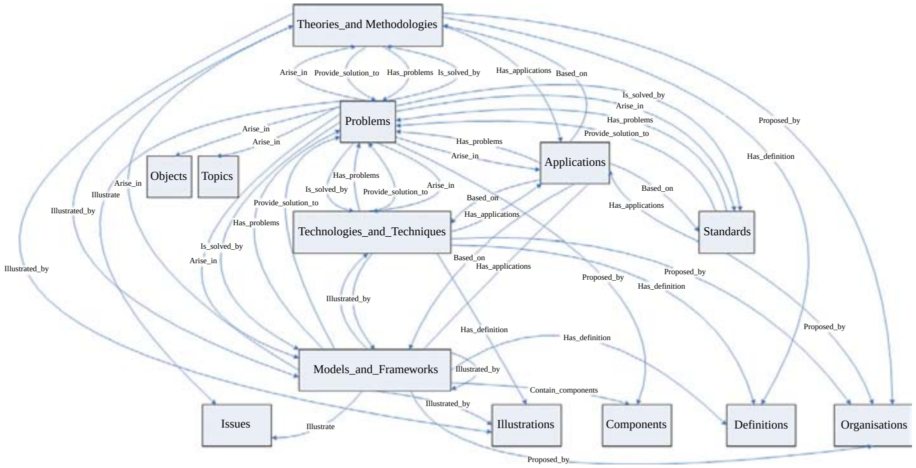


Figure 5.
Some classes and relations in the DL course ontology

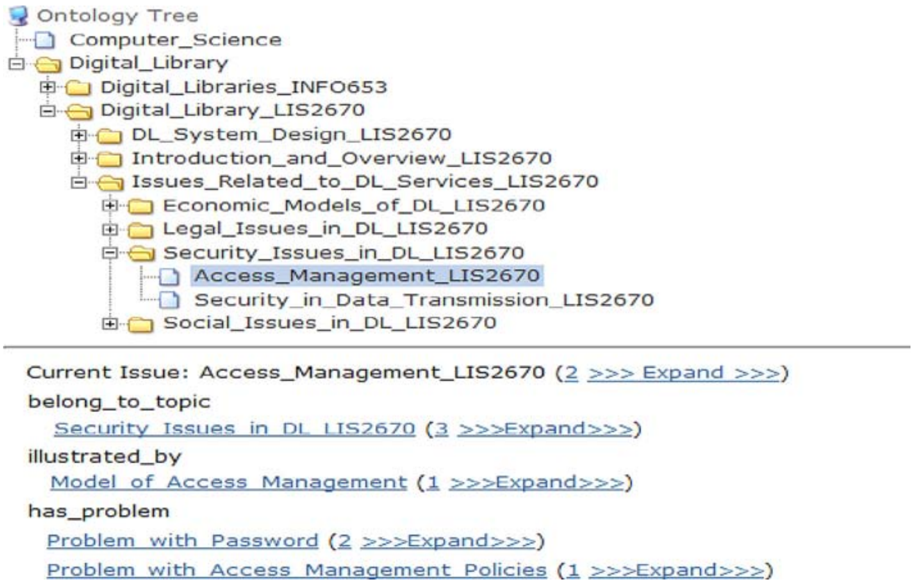
Lecture-based browsing. A course contains a series of lectures. Students are often required to access materials by lectures, or remember course topics by lectures. Therefore, the lectures and a clear association between them and course materials are shown in DiLight. We utilise the "collection" level 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 and their corresponding documents (Figure 6).

Ontology-based browsing. Dicheva and Dichev (2004) stated that students are often unaware of the complete context of a learning task, and thus need help in getting oriented to the conceptual structure of the subject domain. The ontology we constructed can be utilised as a visual presentation of the conceptual structure of DL course topics. With the 43 types of relationships provided in the ontology, students can navigate through the hierarchical structure, and explore or discover internal semantic relationships among course topics/concepts. For example, from the "Access Management" topic it is possible and reasonable for students to jump to the "Security Issues in DL" and "Security in Data Transmission" topics (Figure 7). We acknowledge that it is not ideal to visualise the ontology as a hierarchical tree structure because it artificially imposes certain orders among the topics. However, what motivated us to take this approach is that it is relatively straightforward to integrate the ontology with the rest of the DiLight interface, which is basically a text-based webpage environment. We are working on a java-based visualisation module as a future enhancement of DiLight. In addition to the associated relationships between nodes in the ontology, such as "illustrated by" and "has problems" presented in Figure 7, the visualised ontology is still a network, not a strict tree structure.

Ontology-based search. Similar to searching in other environments, one of the biggest problems that students face in their retrieval of course materials is which terms or keywords should be used in their queries. Simple keyword searches are valuable

⊟Unit 01 - Introduction of Digital Library
  Agenda and Goals
  What is a Library
  What is Digital
  DL Examples
  What is Digital Library Def
  Explanation of DL Definition
  Explanation of What is DL Issue
  Why people work on digital libraries
  Why Librarians work on Digital Libraries
  Fourth International Summer School on Digital Libraries
  Why Computer Scientists Work on Digital Libraries
  Computer Scientists Topics on Digital Libraries
  Why Stakeholders Work on Digital Libraries
  Teachers Topics on Digital Libraries
  Digital Library Faith
  DL and Other Topics
  How DL Becomes Realities
  One Motivating Force and Internet Users
  Questions to Librarians
  Digital Libraries and Traditonal Libraries
  Course Introduction
  Early Visionaries
⊞Unit 02 - Architecture of Digital Library
⊞Unit 03 - Representation of Digital Objects
⊞Unit 04 - Metadata
⊞Unit 05 - XML and Markup Languages

Figure 6.
Lecture-based browsing in DiLight



- Ontology Tree
  - Computer_Science
  - Digital_Library
    - Digital_Libraries_INFO653
    - Digital_Library_LIS2670
      - DL_System_Design_LIS2670
      - Introduction_and_Overview_LIS2670
      - Issues_Related_to_DL_Services_LIS2670
        - Economic_Models_of_DL_LIS2670
        - Legal_Issues_in_DL_LIS2670
        - Security_Issues_in_DL_LIS2670
          - Access_Management_LIS2670
          - Security_in_Data_Transmission_LIS2670
        - Social_Issues_in_DL_LIS2670

Current Issue: Access_Management_LIS2670 (2 >>> Expand >>>)

belong_to_topic

Security Issues in DL LIS2670 (3 >>>Expand>>>)

illustrated_by

Model of Access Management (1 >>>Expand>>>)

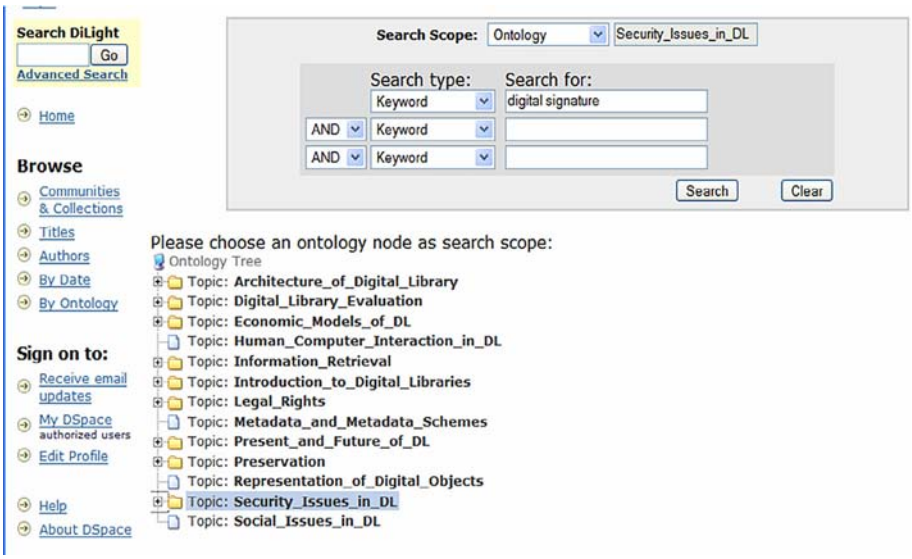
has_problem

Problem with Password (2 >>>Expand>>>)

Problem with Access Management Policies (1 >>>Expand>>>)

Figure 7.
Ontology-based browsing and associated recommendation in DiLight

when students know exactly what they are looking for and/or the information inside documents can be well defined as unique keywords. However, keyword searches often return irrelevant documents containing homonyms (e.g. "organisation" used in collection building and in communities), and miss truly relevant documents that use synonyms (e.g. "architecture" vs "framework") or abbreviations (e.g. "Digital Library" and "DL"). Since teachers and students may have different viewpoints about the materials, and/or different levels of knowledge about the learning materials, the above-mentioned term mismatch problem could be common in e-learning systems. An ontology-based search can overcome this problem by utilising its interconnected concepts and associative relationships. It can locate documents that are semantically similar to the query terms regardless of the surface difference, thus achieving more accurate or comprehensive retrieval results. For instance, our collection contains three documents concerning data encryption. Document A is about Message-Digest Algorithm 5 (MD5), B is about RSA (which stands for Rivest, Shamir and Adleman, a public-key encryption technology), and C is about Data Encryption Standard (DES). A is assigned to ontology node "Algorithm in Data Transmission", whereas B and C are assigned to the 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 because nodes A and C are related to B, even though neither A nor C contains the keyword "RSA". Another advantage of ontology-based search is that it provides facilities to explicitly specify the scope of a search. This is especially useful when the user has knowledge about the possible areas/scopes of the information sought. A clearly defined scope helps to reduce ambiguities in query terms and increase search accuracy (Vallet *et al.*, 2005). For example, as shown in Figure 8, students have been trained to know that if



Search Scope: Ontology Security_Issues_in_DL

Search type: Keyword Search for: digital signature

Please choose an ontology node as search scope:

Topic: Architecture_of_Digital_Library
Topic: Digital_Library_Evaluation
Topic: Economic_Models_of_DL
Topic: Human_Computer_Interaction_in_DL
Topic: Information_Retrieval
Topic: Introduction_to_Digital_Libraries
Topic: Legal_Rights
Topic: Metadata_and_Metadata_Schemes
Topic: Present_and_Future_of_DL
Topic: Preservation
Topic: Representation_of_Digital_Objects
Topic: Security_Issues_in_DL
Topic: Social_Issues_in_DL

Figure 8.
Scope specification in ontology-based search

"digital signature" is within the scope of security issues in DL, they can explicitly identify the search scope as within the topic "Security Issues in DL" instead of searching the whole data repository of 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 to the topics over the whole course. Both of them are limited in providing navigation beyond the immediately linked topics (Large *et al.*, 1999). However, sometimes it is important for students to build up understanding about related topics that are taught in different classes or under different circumstances. When the students' understanding of a topic is low, they need to be directed intelligently towards relevant resources. For example, it is important for students to understand that the discussion of different communities in DL in 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 utilise the "derived from" relationship in our ontology to explicitly markup the connection. With such relationships, documents in both topics are linked together across the ontology. Thus, an associated recommendation is provided when students access either of the topics.

A summative evaluation of DiLight

The ultimate goal of our evaluations is to understand the degree to which DiLight will help students in their studies. Ideally, examinations of students' interactions with DiLight would be conducted in the students' own study settings. However, for our summative evaluation, we took the route of controlled experiments, where the tasks, the systems used, and the order of completion of the tasks on the systems were designed and assigned by us, the experimenters, rather than developed and determined by the students. This level of control helped us to avoid factors that could affect the results but are irrelevant to the goal of the study, so that we could concentrate on the specific research questions to be answered (Grubišić *et al.*, 2009). It also enabled us to collect rich data about the students and their interactions with the systems so that a deep analysis can be performed.

Experiment design

According to Bloom's Taxonomy (Bloom and Krathwohl, 1956), students' learning skills can be divided into six different levels: knowledge, comprehension, application, analysis, synthesis, and evaluation. Students cannot obtain higher level skills unless they have solid foundations for the lower level skills. In our experiment, therefore, we concentrated on the basic level (knowledge) and a middle level (application) with the goal to examine whether DiLight can provide superior support over existing commercial learning support systems.

Our study concentrated on examining the systems' support for accessing materials. Our motivation was that retrieving information is one of the most important aspects of the learning process, and a good learning support system should have the capability to help students find the information they want effectively and efficiently. Therefore, with the assumption that a good learning support system would help students recall and thus retrieve the materials in the collection (i.e. related to the knowledge level of learning skills) and associate different concepts in the collection and locate the materials for confirmation (i.e. related to the application level of learning skills). The hypotheses examined in our formal summative evaluation were:

H1. Students can recall and locate course materials effectively when using the DiLight system.

H2. Students can apply their knowledge and associate several course-related materials effectively when using the DiLight system.

Consequently, the tasks that students performed in our summative evaluation could be classified into two groups. The first group consisted of four tasks (T1-T4 in the following discussion), whose goal was to test *H1*. They concentrated on recalling and locating specific concepts or topics in the collection. For example, T3 asked students to find the slide number of the course topic "what is lossless compression?" The second group, which was to test *H2*, also contained four tasks (T5-T8). They focused on associating various concepts based on the students' knowledge and/or the help of the system. For example, task T5 was about "which techniques can be used to solve the problem of 'which objects are to be digitised'?", and the students should find the slide number of the course material that could be used to support their answers. The details of the eight tasks are as follows:

- *T1*. Find the document that contains the answer to "which layer of the 5s model is related to the study of organising digital objects into collections?".
- *T2*. Find the document that contains the information supporting the statement "URN is related to identifiers of digital objects".
- *T3*. Find the document that contains the answer to "what is lossless compression?".

- *T4*. Find the document that contains the information that is "the illustration of structural metadata".
- *T5*. Find the document(s) that contain the answer to "which techniques can be used to solve the problem of 'which objects are to be digitised'?".
- *T6*. Find the document(s) that contain the answer to "which problem do metadata schema registries try to solve?".
- *T7*. Find the document(s) that contain the answer to "which retrieval scenario has the "deep web" issue?".
- *T8*. Find the document(s) that contain the answer to "digital watermark is a technique to resolve which kind of problems?".

Partially to keep the experiment under a certain time limit, and partially to simulate a situation where students are required to find answers under time pressure such as in exams, we set a five-minute time limit for each task to be completed (Zaiane and Luo, 2001). We acknowledge that this excludes access scenarios where students do not have any time limit. However, we think that there is adequate justification to make the time limit from both the experiment procedure and the efficiency of the e-learning systems.

Our experiment is comparative in nature so the effectiveness of our system can be demonstrated by comparing it with a baseline system (Montgomery, 2006). The baseline system used in the experiment was CourseWeb (essentially the BlackBoard system). It has been widely used in various teaching and learning environments, and students in the DL course have been using it since the beginning of the course. Therefore, the participants were very familiar with the system, including the organisation of the course-related materials in it.

Our experiment took the commonly used within-subject design (Greenwald, 1976) where each student performed repeated trials (e.g. several accesses, each for a different task) on both the DiLight system and the baseline system. The order of those trials was varied systematically in order to block (i.e. average out) the effect of presentation order on learning and fatigue and the effects of individual differences in students and tasks because of different levels of task difficulty. Consequently, we can be confident that our analysis focuses on the desired effect factor (i.e. the systems). A Latin Square was used as the basis for rotating presentation order.

The data used in the experiment include all slides presented in the classes, a total of 988 slides. Since only half of the course had been taught at the time of the experiment, materials from the first six weeks – 585 slides in total – were used for the experiment. The eight tasks mentioned above were based on this collection.

We adopted several performance measures from the literature in this experiment (Mark and Greer, 1993). The first one is task completion rate (TCR), which is the percentage of the participants who correctly found the right answer for a specific topic. In our collection, the correct answer to a task might appear in several places. This is because a topic might be discussed more thoroughly when it is first introduced at one point in the course, but it could also be briefly repeated or reviewed later in the course. Although all those mentions could be the correct answers, it is much more preferable to find the first primary mention that was discussed in the class. Therefore, we differentiate a strict version of task completion rate (i.e. strict completion rate, STCR) which only counts the finding of the first primary mention as the correct answer, and a loose version of task completion rate (i.e. loose completion rate, LTCR), where finding any mention of the answer is viewed as task completed. Both of the measures were used in the result analysis.

The second measure is task completion speed (TCS) which is the time that a student used to finish a task. If a task was not finished before the end of the allocated time (i.e. five minutes), the maximum allocation time (i.e. five minutes) was used as the time. If the answer that the participant found was wrong, the task was treated as incomplete, and thus five minutes was taken as the completion time rather than their raw time.

The formal experiment for each participant lasted for about 90 minutes. The experiment always started with a 15 minutes of training to let students refresh their memory about using various tools provided by CourseWeb and DiLight. They then performed four tasks using one system, and switched to the other system to perform the remaining four tasks. Tasks regarding *H1* and *H2* were mixed together when they were plotted in a rotation matrix based on the Latin Square principle (Montgomery, 2006). There was an entry questionnaire collecting students' basic demographic data including their search and general DL background. A pre-task questionnaire was presented before each task to determine the student's familiarity with that specific task. Students were also asked to fill in a post-task questionnaire after finishing each task and a post-system questionnaire after having done all four tasks on one system.

Participant profiles

A total of ten students participated in our summative evaluation. Their profiles are:

- Most (seven) participants were native speakers of English.
- Most (seven) participants were female, and the average age of all participants was 28, with the youngest being 24 and the oldest being 47.
- All participants reported that they use computers and search services very frequently. Eight participants reported that they had knowledge about basic IT technology such as CPU, TCP/IP, and HTML.
- All participants had search experience. The average duration of their search experience was about seven years, with a minimum of two years and maximum of ten. Their average local library search experience was good. Their average commercial online system search experience was moderate. Their average www-search experience was excellent.
- All participants were LIS students, and had moderate experience with DL.
- All participants had experience in using the CourseWeb system.

Results analysis and discussion

Analysis by TCR

When examining the results in terms of TCR, we found that participants performed better when they used DiLight than when they used CourseWeb under both LCR (0.875 vs 0.55) and SCR (0.725 vs 0.4). Such differences are statistically significant in both cases (paired two sample one tail *t*-test, $p = 0.03$). This demonstrates that DiLight is useful in helping students to complete their learning tasks.

However, from Figures 9 and 10, we can see that the statistical differences all come from tasks related to *H2* (T5-T8); students' performance did not differ much between the two systems in their *H1* related tasks (T1-T4), especially under the measure of STCR.

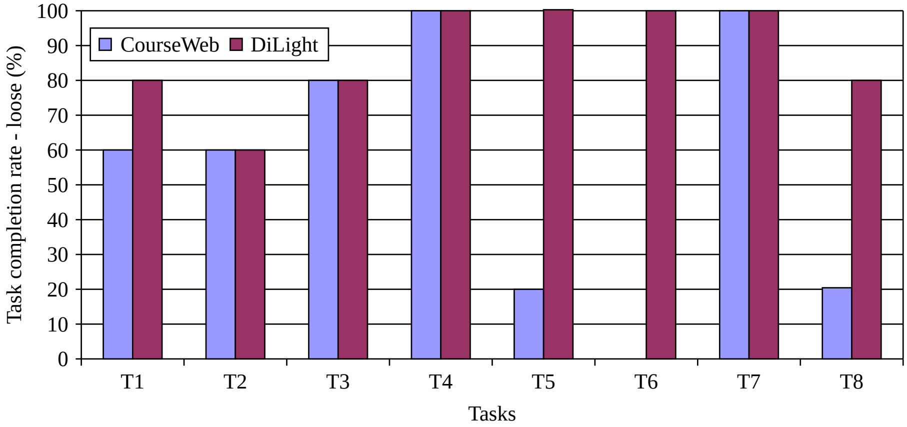


Figure 9.
The TCR (loose) between the two systems

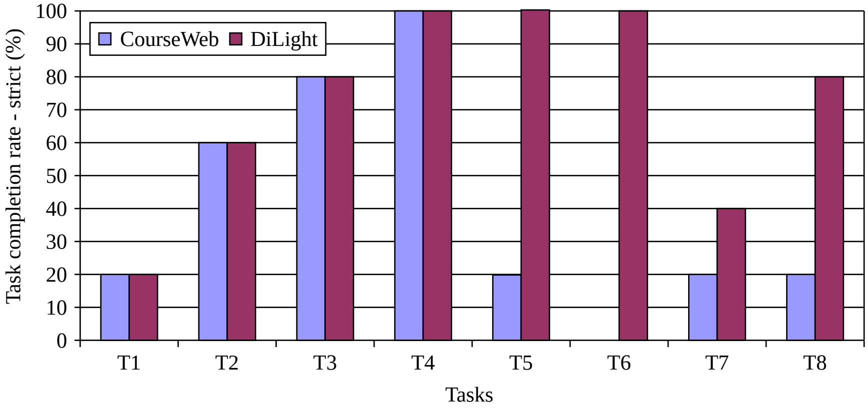


Figure 10.
The TCR (strict) between the two systems

This seems to indicate that both systems provide similar support for subject recall and retrieval of course topics for simple tasks. In this situation, the task difference is actually a much bigger factor than the system difference.

Task T1 is an anomaly as shown in Figures 9 and 10. This task asked subjects to find the answer to the question "which layer of the 5s model is related to the study of organising digital objects into collections?" Although participants using DiLight performed better than those using CourseWeb when measured by LCR, there was no difference at all when measured by SCR, and there was a drop between the values of LCR and SCR in results from both systems. It seems that participants managed to find some related answers, but not the primary answer we wanted. Interestingly, subjects using DiLight reported lower familiarity (1.8 vs 2.8) and higher difficulty (3.8 vs 3.2) with the task than those using CourseWeb, but they managed to perform better when measured by LCR.

Figures 9 and 10 also show that the significant difference between the two systems comes from the performance of the group of tasks related to *H2*. On average, the difference between using DiLight and using CourseWeb was 0.35 vs 0.95 when measured by LCR, and 0.15 vs 0.8 when measured by SCR. The very low TCRs when using CourseWeb under both measures indicate that participants were struggling and probably did not receive much help from the system in this set of tasks. In contrast, subjects using DiLight still performed very well even though the tasks had become much more difficult.

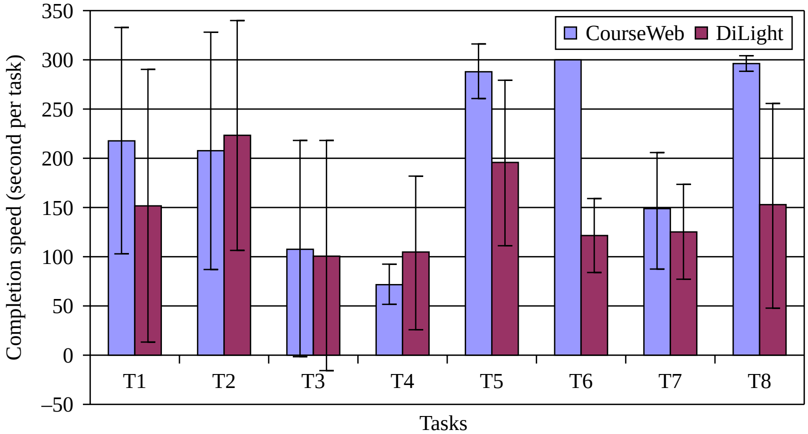
T6 is an interesting task. It asked students to find the problem that "metadata schema registries" try to solve. All subjects using DiLight found the correct answers. This is why both the SCR and LCR are 100 percent. However, all subjects using CourseWeb failed at this task. Our further examination of the data showed that among the five students who performed T6 using CourseWeb, two students found wrong answers within five minutes, and the other three did not find any answer at all. Both groups of students indicated that they had relatively low familiarity with the task (both at 1.8), but the DiLight group gave much lower ratings of task difficulty than the CourseWeb group (2.2 vs 4.4). All these results seem to indicate that students received good quality support when using DiLight for this task.

Analysis by TCS

Overall, compared with using CourseWeb, students spent less time using DiLight to complete their tasks (see Figure 11 measured by TCS). The average was 146 seconds vs 204 seconds per task. The difference is statistically significant (paired two sample one-tail t-test, $p = 0.034$). The difference between the two systems also shows that samples from DiLight had smaller standard deviation than that of CourseWeb (96 vs 108), which indicates that students using DiLight performed more consistently than those using CourseWeb.

In Figure 11, we can see that students using DiLight took less time on T5 to T8 (the group of tasks related to *H2*) than those using CourseWeb. This is consistent with what we have observed in the analysis by TCR.

However, we could not see such clear superiority when examining the TCS of using DiLight compared with using CourseWeb on T1 to T4 (the group related to *H1*).



Note: The Y-error bars indicate the range of the values

Figure 11.
TCS between the two systems, arranged by tasks

Although overall we still can see that using DiLight can help students complete tasks faster than using CourseWeb, but the difference is much smaller than for T5 to T8. We can see that it actually took longer to complete T2 and T4 using DiLight.

T4 asked students to find an illustration of structural metadata presented in the class. This task was rated as the most familiar task among all eight tasks, and students using DiLight expressed slightly higher familiarity (3.8 vs 3.2). This task was also considered the easiest task by students using CourseWeb (i.e. 1.6), and the second easiest task by those using DiLight (i.e. 2.2). All these ratings were reflected by the 100 percent TCR and the fastest TCS (in the case of CourseWeb) and near fastest (in the case of DiLight) by both groups of students.

However, we also notice that the TCS using DiLight is not as good on T4 as that of using CourseWeb, unlike the other tasks. We have an untested hypothesis about the reason. We know that students can only use browsing as the means to access materials in CourseWeb, and we also know from students' feedback that they liked using the various search methods provided in DiLight to find answers for their tasks. Therefore, the advantages and limitations of browsing against search would affect their performance between the two systems (Mackinlay and Zellweger, 1995). T4 is a task that might be more difficult to complete via search. This is because the two important terms related to T4 are "illustration" and "structural metadata". The latter can be used as a good keyword for search, but the first one is a descriptive term which is not a good search term. This is because it only indicates that students need to find a diagram or graphic presentation of structural metadata, which may not be accompanied by the word "illustration". Since browsing would not be affected by this problem, we think that this is the reason that students performed slower when using DiLight. Of course, browsing is provided in DiLight too, but it is usually used less often.

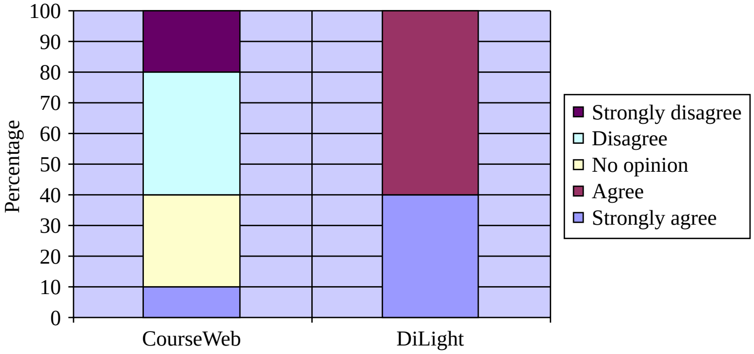
Analysis by subjective views

In the questionnaires, we gave to students after using each system, we asked for their feedback on the support of the two systems and on the usefulness of the main access methods provided by the two systems. The two questions were:

(1) the system you just used helpful in supporting your task?
(2) the specific access methods in the system helpful in supporting your task?

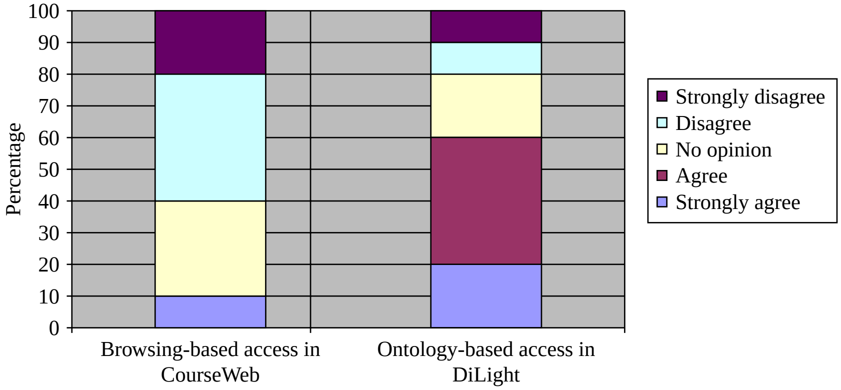
Figure 12 shows that all the students agreed or strongly agreed that DiLight was helpful in supporting their tasks. This was consistent with what we obtained from the students' performance on the tasks. However, only 10 percent of the students strongly agreed that CourseWeb was helpful. At the same time, 60 percent of students actually disagreed or strongly disagreed that CourseWeb was helpful. We did not collect such information after each task, so we could not tell whether students' opinions on CourseWeb would differ between the two groups of tasks.

Figure 13 shows that 60 percent of the students agreed or strongly agreed that the ontology-based access methods in DiLight were helpful in supporting them to find answers. Only 20 percent of students disagreed or strongly disagreed with that statement. However, only 10 percent of students who used CourseWeb strongly agreed that the browsing-based access method in CourseWeb was helpful. On the contrary, about 60 percent of students disagreed or strongly disagreed with that statement.



Note: "Was the system helpful in supporting your task?"

Figure 12.
Students' response to the question



Note: "Was the specific access method in the system very helpful in supporting your task?"

Figure 13.
Students' responses to the question

Discussion of the findings

The results from the summative evaluation show that DiLight can provide significantly better support for students' learning tasks than that of CourseWeb, one of the widely used commercial e-learning systems. Therefore, DiLight has demonstrated its usefulness in e-learning environments.

As to the two hypotheses studied in the formal summative evaluation, we have to reject *H1* because both systems provided some support for students' basic knowledge recall and retrieval from the collection. Although the results from DiLight were slightly better in terms of both TCR and TCS, the difference was not significant. However, we did find that DiLight helped students better even when students were either less familiar with the tasks or felt the tasks were more difficult.

We accept *H2* because students using DiLight consistently outperformed those using CourseWeb on those tasks related to *H2*. We can see that it was because of the performance on these tasks that the DiLight system had significantly better results than CourseWeb. This demonstrates that DiLight can help students in applying what they have learned into tasks involving relationship discovery and problem solving.

Students' performance on T4 might demonstrate that there is no single access method that is the best for all situations. Sometimes, search-based methods can find the answers quickly but in other cases browsing-based methods may be better. This further confirms our belief that there should be multiple IA methods built into e-learning systems to cope with different tasks, demands or preferences from students, and other possible situations. Of course, we acknowledge that more studies are needed to further confirm this finding.

The experiment demonstrated that DiLight has high quality in functionality and usability. It contains almost all the information that is needed for the DL course, including readings, slides, and videos. Browsing, search and ontology-based functions are adequate for students to accomplish their tasks. Although the search function (including basic search and advanced search) was generally the first choice of the students, ontology-based methods were useful in supporting their tasks too.

Of course, the current implementation of DiLight still has its limitations. To input course materials into the system, the teacher has to manually construct the configure file so that the slides and other related materials can be correctly divided into topic based documents. For motivated teachers this would not be an issue since the configure file can be a by-product of their design of the course materials. However, when the amount of materials to be added is large, it is time consuming. One possible extension of DiLight to solve this limitation is to use an automatic clustering technique to group slides that are similar in content together. Another limitation is related to the ontology building for the course content. Right now, the ontology related to the course materials is built manually by the teachers. Again, for motivated teachers who know the topic area well, with modern ontology building tools, this is not an impossible task. More importantly, along with the further development of knowledge representation in many topic areas, ontology is becoming available for many domains. Such ontology can be easily integrated into DiLight to provide ontology based IA. No matter how these two limitations are addressed, the basic innovation features of DiLight would not be affected.

Conclusion and future work

With the advance of web-based information technology, e-learning technology has become an important extension to traditional instruction methods. However, it also faces some challenges in terms of collection management, material access methods, and collaboration among students and teachers. We believe that DL technologies, combined with knowledge representation and IA technologies, can be the underlying platform for developing an effective e-learning system. DiLight is such a system and was developed to provide interactive, integrated, and effective learning support for students. Using DSpace as the framework, DiLight provides powerful and flexible collection management capabilities, knowledge-rich content representation, and multiple IA methods including ontology-based methods to help students with different backgrounds, needs, and preferences.

We also presented the evaluations of the usefulness of the DiLight system. Through a formal summative experiment, we have examined the roles that DiLight could play in supporting students' learning tasks in the form of basic knowledge recall and retrieval, and more advanced knowledge association. CourseWeb, a local version of BlackBoard, was used as the baseline in the formal evaluation. The experiment results show that DiLight can provide slightly better support than the baseline in helping students' basic knowledge recall and retrieval tasks, but DiLight's support was significantly better in helping students' advanced knowledge association tasks. The multiple access methods were considered the major contribution of this system, and a large majority of students praised the usefulness of DiLight. Our study also shows that it is necessary to provide multiple access methods in learning support systems, because students may need different tools to access materials, depending on the task. The study also shows that, although the ontology is a useful presentation of the topics in the DL course, students need more time and mental effort to master the concept of ontology and to develop strategies for using it sensibly.

Further directions of this work include examining the usefulness and robustness of DiLight with more courses, integrating ontology from different sources to express different views of the topics, and developing collaborative mechanisms to support learning and tutoring among students.

References

Acosta-Diaz, R., Guillen, H.M., GarciaRuiz, M.A., Gallardo, A.R., Pulido, J.R.G. and Reyes, P.D. (2006), "An open source platform for indexing and retrieval of multimedia information from a digital library of graduate thesis", *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education E-Learning, Association for the Advancement of Computers in Education, Chesapeake, VA*, pp. 1822-9.

Ahn, J.-W., Brusilovsky, P. and Sonsnovsky, S. (2006), "Quizvibe: accessing educational objects with adaptive relevance-based visualization", *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education E-Learning, AACE, Chesapeake, VA*, pp. 2707-14.

Alexander, S. and McKenzie, J. (1998), *An Evaluation of Information Technology Projects in University Learning*, Department of Employment, Education and Training and Youth Affairs, Australian Government Publishing Services, Canberrra.

Bloom, B.S. and Krathwohl, D.R. (1956), *Taxonomy of Educational Objectives: The Classification of Educational Goals Handbook I: Cognitive Domain*, Longmans, Green, NY.

Borgman, C. (2000), *From Gutenberg to Global Information Infrastructure*, MIT Press, Boston, MA.

Brown, W.E., Lovett, M., Bajzek, D.M. and Burnette, J.M. (2006), "Improving the feedback cycle to improve learning in introductory biology using the digital dashboard", *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education, AACE, Chesapeake, VA*, pp. 1030-5.

Brusilovsky, P. (1999), "Adaptive and intelligent technologies for web-based education", *Kunstliche Intelligenz*, Vol. 4, pp. 19-25 (Special Issue on Intelligent Systems and Teleteaching).

Brusilovsky, P. and Peylo, C. (2003), "Adaptive and intelligent web-based educational systems", *International Journal of Artificial Intelligence in Education*, Vol. 13 Nos 2-4, pp. 159-72.

Burgess, J.R.D. and Russell, J.E.A. (2003), "The effectiveness of distance learning initiatives in organizations", *Journal of Vocational Behavior*, Vol. 63 No. 2, pp. 289-303.

Carlson, L.E. and Sullivan, J.F. (1999), "Hands-on engineering: learning by doing in the integrated teaching and learning program", *International Journal on Engineering Education*, Vol. 15 No. 1, pp. 20-31.

Conlan, O., Hampson, C., O'Keeffe, I. and Heller, J. (2006), "Using knowledge space theory to support learner modeling and personalization", *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education, AACE, Chesapeake, VA*, pp. 1912-9.

DeRouin, R.E., Fritzsche, B.A. and Salas, E. (2005), "E-learning in organizations", *Journal of Management*, Vol. 31 No. 6, pp. 920-40.

Dicheva, D. and Dichev, C. (2004), "A framework for concept-based digital course libraries", *Journal of Interactive Learning Research*, Vol. 15 No. 4, pp. 347-64.

Falquet, G. and Ziswiler, J.C. (2005), "A virtual hyperbooks model to support collaborative learning", *International Journal on E-Learning*, Vol. 4 No. 1, pp. 39-56.

Galagan, P.A. (2000), "The e-learning revolution", *Training & Development*, Vol. 54 No. 12, pp. 24-30.

Greenwald, A.G. (1976), "Within-subjects designs: to use or not to use?", *Psychological Bulletin*, Vol. 83 No. 2, pp. 314-20.

Gruber, T. (1993), "A translation approach to portable ontology specifications", *Knowledge Acquisition*, Vol. 5 No. 2, pp. 199-220.

Grubišić, A., Stankov, S., Rosić, M. and Itko, B. (2009), "Controlled experiment replication in evaluation of e-learning system's educational influence", *Computers and Education*, Vol. 53 No. 3, pp. 591-602.

Gunasekaran, A., McNeil, R.D. and Shaul, D. (2002), "E-learning: research and applications", *Industrial & Commercial Training*, Vol. 34 No. 2, pp. 44-53.

Henry, P. (2001), "E-learning technology, content, and services", *Education + Training*, Vol. 43 No. 4, pp. 249-55.

Henze, N., Dolog, P. and Nejdl, W. (2004), "Reasoning and ontologies for personalized e-learning", *Educational Technology & Society*, Vol. 7 No. 4, pp. 70-81 (Special Issue on Ontologies and the Semantic Web for E-learning).

Horrocks, I. and Hendler, J. (Eds) (2002), *Proceedings of the First International Semantic Web Conference*, Springer, Berlin.

Horton, W. (2000), *Designing Web-based Training*, Wiley, New York, NY.

Kosarzycki, M.P., Salas, E., Derouin, R. and Fiore, S.M. (2003), "Distance learning in organizations: a review and assessment of future needs", Salas, E. (Series Ed.) and Stone, D. (Vol. Ed.), *Advances in Human Performance and Cognitive Engineering Research: Vol. 3. Human Resources Technology*, JAI, Boston, MA, pp. 69-98.

Large, A., Tedd, L.A. and Hartley, R.J. (1999), *Information Seeking in the Online Age: Principles and Practice*, Bowker Saur, Munich.

Mackinlay, J.D. and Zellweger, P.T. (1995), "Browsing vs. search: can we find a synergy?", in Katz, I., Mack, R. and Marks, L. (Eds), *Conference Companion on Human Factors in Computing Systems*, ACM, New York, NY, pp. 179-80.

Mark, M.A. and Greer, J.E. (1993), "Evaluation methodologies for intelligent tutoring systems", *Journal of Artificial Intelligence in Education*, Vol. 4 Nos 2/3, pp. 129-53.

Mertens, R., Farxan, R. and Brusilovsky, P. (2006), "Social navigation in web lectures", in Wiil, U.K., Nurnberg, P.J. and Rubart, J. (Eds), *Proceedings of the Seventeenth ACM Conference on Hypertext and Hypermedia*, ACM Press, New York, NY, pp. 41-4.

Meyers, C. and Jones, T.B. (1993), *Promoting Active Learning: Strategies for the College Classroom*, Jossey-Bass, San Francisco, CA.

Minka, T.P. and Picard, R.W. (1996), "Interactive learning with a 'society of models'", *Conference on Computer Vision and Pattern Recognition*, IEEE, Washington, DC, p. 447.

Montgomery, D.C. (2006), *Design and Analysis of Experiments*, Wiley, Hoboken, NJ.

Mwanza, D. and Engestrom, Y. (2003), "Pedagogical adeptness in the design of e-learning environments: experience from the lab@future project", *Proceedings of the World*

Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education E-Learning, AACE, Chesapeake, VA, pp. 1344-7.

Nokelainen, P., Tirri, H., Kurhila, J. and Miettinen, M. (2002), "Use of social navigation features in collaborative e-learning", in Richards, G. (Ed.), *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, AACE, Chesapeake, VA*, pp. 1738-41.

Noy, N. and McGuinness, D. (2001), "Ontology development 101: a guide to creating your first ontology", Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, available at: www-ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf (accessed 8 June 2010).

Paradigm, T. (2001), "Course participation: an active learning approach employing student documentation", *Journal of Marketing Education*, Vol. 23 No. 3, pp. 187-94.

Parker, L.M.P. and Johnson, R.E. (1990), "Does order of presentation affect users' judgement of documents?", *Journal of the American Society for Information Science*, Vol. 41 No. 7, pp. 493-4.

Rosenberg, M.J. (2001), *E-learning: Strategies for Delivering Knowledge in the Digital Age*, McGraw-Hill, New York, NY.

Salas, E., Kosarzycki, M.P., Burke, C.S., Fiore, S.M. and Stone, D.L. (2002), "Emerging themes in distance learning research and practice: some food for thought", *International Journal of Management Reviews*, Vol. 4 No. 2, pp. 135-53.

Saracevic, T. and Dalbello, M. (2001), "A survey of digital library education", *Proceedings of Annual Meeting of American Society for Information Science and Technology*, ASIST, Silver Spring, MD, pp. 209-23.

Shum, S.B. (1998), "Evolving the web for scientific knowledge: first steps towards an 'HCI knowledge web'", *Interfaces Magazine*, Vol. 39, pp. 16-21.

Siritongthaworn, S., Krairit, D., Dimmitt, N. and Paul, H. (2006), "The study of e-learning technology implementation: a preliminary investigation of universities in Thailand", *Education Information Technology*, Vol. 11 No. 2, pp. 137-60.

Spector, J.M. (2001), "An overview of progress and problems in educational technology", *Interactive Educational Multimedia*, No. 3, pp. 27-37.

Spink, A. and Cool, C. (1999), "Education for digital libraries", *D-Lib Magazine*, Vol. 5 No. 5, available at: www.dlib.org/dlib/may99/05spink.html (accessed 8 June 2010).

Studer, R. and Stabb, S. (2003), *Handbook on Ontologies in Information Systems*, Springer, Berlin.

Tane, J., Schmitz, C., Stumme, G., Staab, S. and Studer, R. (2003), "The courseware watchdog: an ontology-based tool for finding and organizing learning material", *Fachtagung Mobiles Lernen und Forschen*, Kassel University Press, Kassel, pp. 93-104.

Vallet, D., Fernandez, M. and Castells, P. (2005), "An ontology-based information retrieval model", in Gómez-Pérez, A. and Euzenat, J. (Eds), *The Semantic Web: Research and Applications*, Springer, Berlin, pp. 455-70.

Vergara, M., Capuano, N. and Sangineto, E. (2003), "Diogene: a training web broker for ICT professionals", *Proceedings of the 2nd International Conference on Multimedia and ICT in Education, Badajoz, Spain*, pp. 330-4.

Welsh, E.T., Wanberg, C.R., Brown, K.G. and Simmering, M.J. (2003), "E-learning: emerging uses, empirical results and future directions", *International Journal of Training & Development*, Vol. 7 No. 4, pp. 245-58.

Zaiane, O.R. and Luo, J. (2001), "Towards evaluating learners' behaviour in a web-based distance learning environment", *The Second IEEE International Conference on Advanced Learning Technologies* (*ICALT'01*)*, Madison, Wisconsin*, IEEE, Washington, DC, pp. 357-60.

Zia, L.L. (2001), "The NSF national science, technology, engineering, and mathematics education digital library (NSDL) program", *D-Lib Magazine*, Vol. 7 No. 11, available at: www.dlib.org/dlib/november01/zia/11zia.html (accessed 8 June 2010).

About the authors

Daqing He is Assistant Professor in the Graduate Program of library and information science at the University of Pittsburgh. His research interests focus on information retrieval, digital libraries, and intelligent information systems. Daqing He is the corresponding author and can be contacted at: dah44@pitt.edu

Yefei Peng is a PhD student in the School of Information Sciences, University of Pittsburgh, and a scientist at Yahoo! Labs. He received a BEng in Engineering Physics from Tsinghua University in 1999 and a MS in nuclear engineering from Purdue University in 2004. His research interests include neural networks, information retrieval, data mining and machine learning.

Ming Mao is a Senior Research Scientist at SAP Labs. She obtained her PhD in information sciences from the University of Pittsburgh. Her research interests include ontology mapping, semantic search and text analysis in business environments. This work was conducted when she was a PhD student in the School of Information Sciences at the University of Pittsburgh.

Dan Wu is Assistant Professor at the School of Information Management, Wuhan University. She obtained her PhD in information science from Peking University. Her research interests focus on information retrieval, information organisation, multilingual information processing, metadata, and digital libraries.