TREE: A Digital Library for Engineering Education

Sherif Elmeligy Abdelhamid, Kenneth J. Reid
Virginia Polytechnic Institute and State University, sherief@vt.edu, kenreid@vt.edu

Abstract - **In this paper we propose TREE, a concept for a new digital library system. Digital libraries are considered as collections of services that operate on digital objects which might include text, visual, audio and video material. According to some researchers, a digital library can be seen as an organization, which might be virtual, that collects, stores and maintains digital content for long periods of time, and provides to its users specialized services of measurable quality according to standard policies.**

In order to understand how to efficiently use digital libraries to support engineering education, an analysis of literature is presented; the results of this analysis are used to identify the main components of our proposed digital library design. The 5S (streams, structures, spaces, scenarios, and societies) framework is used to formally define our digital library. As a follow-up study, the results of this work will be used to guide the implementation and testing of the new system. Additional experiments that involve human subjects will be considered to monitor, evaluate and assess the usage of the system by different users including students, faculty and researchers.

Index Terms – Engineering Education, Digital Library, E-Science, Database Systems, Grid Computing, Workflow Systems

INTRODUCTION

I. Motivation

While there is a large amount of information on the internet related to engineering education, it can be difficult for students, faculty and researchers to explore. The need to handle this problem was our primary motivation to propose the concept of TREE, to provide users with coherent access to a large, organized repository of information related to engineering education.

TREE can provide users with advantages that include: ease of use through user-friendly interfaces, 24/7 access to various resources, and the ability to allow many users to use the same resources at the same time. Searching information via TREE is much easier using keywords or browsing. No space limitation problem exists in contrast to traditional libraries that need more space to physically store library items [1]. TREE will provide support for the engineering education community by giving students, faculty, and researchers an access to wide range of online resources including courses, tutorials, multimedia collections, e-books and biographies. TREE will also provide different services that promote research collaboration amongst researchers, in addition to an online archive for research publications including journals, conference papers, dissertations and theses.

In 1989, a computer scientist Tim Berners-Lee [2] worked for six months at CERN "European Organization for Nuclear Research", he developed an early prototype for the World Wide Web that allowed scientists to share data and communicate much more easily. Now rather than just sharing data, 'the Grid' will help to share computer power. The term Grid comes from a comparison with the electricity grid, a user plugging into the network will access all its resources, without a need to know where the resources are or who owns them. The Grid can be a data Grid, where large amount of data can be stored in a distributed manner, on different nodes. The Grid can also be called a computing Grid, if complex calculations are distributed across many computers to perform faster.

Digital libraries can be seen as systems (possibly distributed) that provide much more services and capabilities than a repository of literature [3]. TREE is designed to support immense data sets that require grid computing. Providing support for e-science will increase the development and achievement in research, through the distributed collaboration of different types of users including researchers, students and teachers. Users will have access to huge computing power provided by the Grid and will be located at distant sites from each other.

II. Challenges

A lot of challenges will be facing TREE, among these will be managing the huge amount of data stored, updated or retrieved by users. Other challenges will be managing the software and processing units, ensuring that the digital library components will act as one unit with no conflict, providing seamless flow through data life cycle, and finally enhancing the scholarly communication among research communities.

RELATED WORK

Many scientists made use of the latest innovations in Computer Science to enhance and advance their research experience. The major turning point was in 1989, when Tim Berners-Lee developed Enquire [4]. Enquire was a hypertext system that allowed sharing of information among physicists. Before Enquire, there were other systems like MEMEX [5] and NLS [6]. The problem with these two systems was the incompatibility among networks, hardware and data types, which make the transfer of information very difficult. In the following subsections we conducted a literature survey within three related topics to our proposed system.

I. Digital Libraries

The role of Digital Libraries in the Education process has been obvious during the past years. Fox, et al. [7] introduced an Educational Digital Library named Ensemble. Ensemble came with two main objectives: ease of delivery of tools and contents, and increasing the interactions and social bonds between computing educators by using forums and posts. Another Example of Educational Digital Library is AlgoViz [8]. AlgoViz is a gateway for online visualizations and animations. It was an open community for users and developers. SMET [9] is another example of Educational Digital Library. Several organizations have established online resources for Engineering Education including WEPAN [10] which focus on research and best practices to promote the inclusion of women in the field of engineering. Another organization is Northwestern Center for Engineering Education (NCEER) [11], which provides a mature infrastructure that promotes collaboration and research. Figure 1, shows how NCEER makes connections between the many different programs and centers across Northwestern, leading to a highly interdisciplinary environment.

II. Knowledge Dissemination and Publishing

Effort has been directed toward accelerating the research cycle. One prominent example was the High Energy Physics archive (HEP), which has about 220,000 articles and more than 12,000 visitors per day. Weatherley [12] proposed a new approach to facilitate the reviewing task of some complex resources, by dividing up the task among community volunteers. The model was named "The Partnership Review Model". Pandey [13] described the Digital Library as a network of resources, for learning, teaching and archiving. Carr, et al., [14] showed the role of Digital Libraries in the Scientific Life Cycle, and proved that the dissemination of knowledge is as important as research itself.

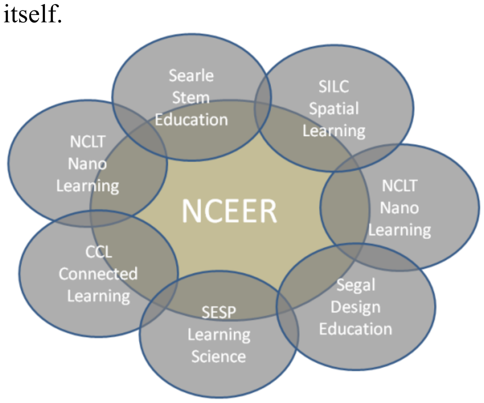

FIGURE 1
NCEER PROGRAMS AND CENTERS [11]

III. Workflow Management Systems

Workflows allow scientists to provide a high-level declarative way of specifying in silico analyses. Researchers can easily describe steps of each experiment as workflow diagrams. Sharing and reusing of workflows can stimulate the pace of research and the overall productivity of experimentation through evident savings in time and effort. In this paper we are interested in scientific data analysis workflows.

Several Workflow Management Systems have emerged during the last ten years, including Triana [15], Kepler [16] and Taverna [17]. Research efforts tried to extend the Workflow Management Systems to the World Wide Web, like Biowep [18]. However Biowep was dedicated only for biological sciences and didn't allow the creation of workflows, just selection and execution. Another Web-based system was BioWMS [19]. At the University of Rochester, a team of researchers developed DEDiscover [20], which was a Workflow-based Differential Equation Modeling Software. The intended users were scientists, statisticians and modelers.

PROPOSED DESIGN

The overall architecture of TREE takes the form of centric layers as shown in figure 2. The outer layer represents the user interface. The second layer is composed of three digital libraries: Workflow Digital Library (WFDL), Publication Digital Library (PubDL), and Education Digital Library (EduDL). The third layer, which represents the core, will be the grid, and will handle all the data storage and processing required by the three digital libraries.

The main goal is to make the three different Digital Libraries act as one; this goal could be achieved using the Open Digital Library Architecture, described in figure 3. The Open Archive Initiative (OAI) protocol [21] will connect together the digital libraries of TREE, where each one will act as a data and service provider. The three layers of TREE will be discussed in detail in the following subsections.

I. User Interface (UI)

Different types of users are going to use TREE, so the user interface will have to provide different views and capabilities to support its users' information seeking needs. The user interface is composed of seven different types of software agents as shown in figure 4. The user interface agent is responsible to handle interaction between the user and the system. The Access control will be the responsibility of the Access control agent. Certain users will have access only to certain views based on their access rights. The broker agent is called the central agent, as it coordinates the flow of information between other agents and the user interface.

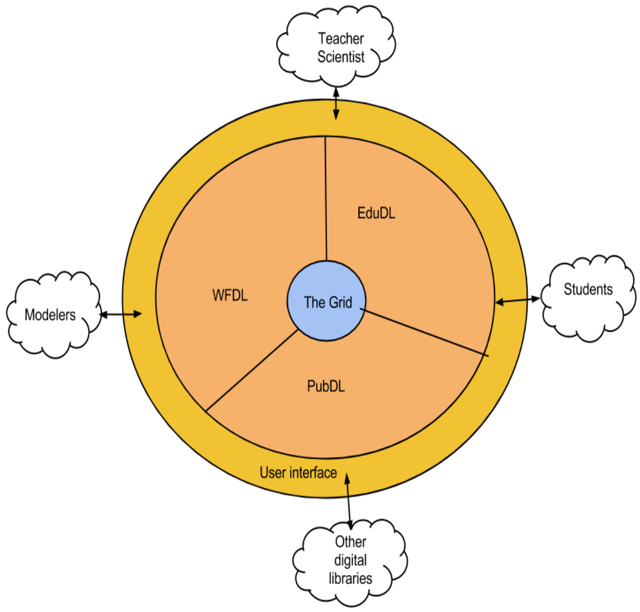

FIGURE 2
DIGITAL LIBRARY DESIGN

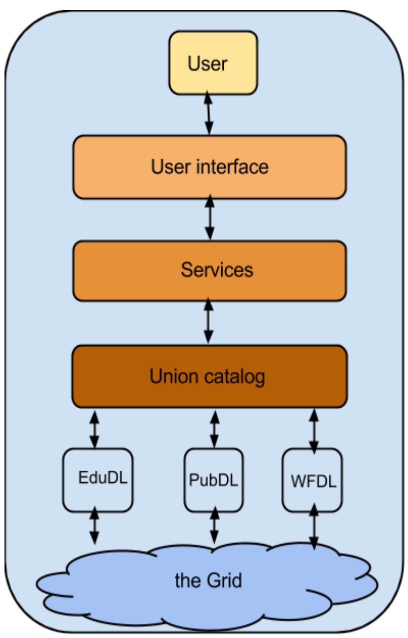

FIGURE 3
OPEN DIGITAL LIBRARY ARCHIVE

The user interface will allow users to use any search term (word, phrase, title, name, subject, etc.) to search the entire collection, and will be able to provide features like borrowing, returning and searching for items. Other features for special users include inventory management. The UI will be designed to support users with different levels of computer skills, and special support will be provided to users with disabilities by providing larger fonts and audio support.

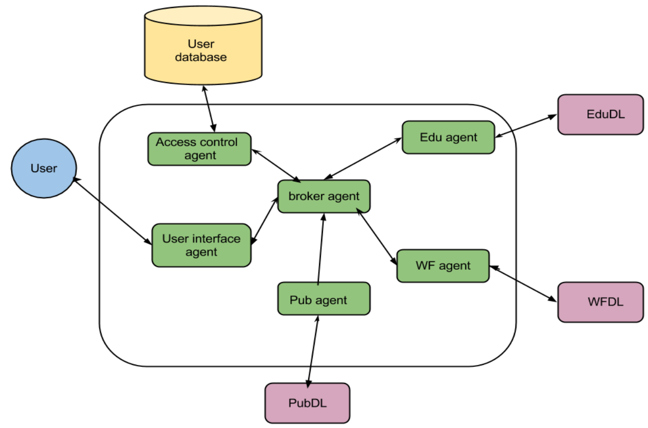

FIGURE 4
USER INTERFACE

II. *Publication Digital Library (PubDL)*

The PubDL will maintain all the information on engineering education publications, including conference papers, posters and journals. PubDL is intended to speed up the research cycle by promoting collaboration among researchers and sharing of experiments' results. Figure 5 shows the PubDL components.

The experiment manager (EM) is responsible for creating, updating and deleting experiments. The experiments metadata is stored in the experiments repository. Metadata includes: researcher name, co-researchers, experiments category, goals, procedures, annotations and results. The EM can also search for existing datasets to be used for testing the experiment. The experiment procedures are described using workflow diagrams.

The data analyst (DA) is a suite of tools created by software developers to help researchers analyze their data using several statistical methods. The DA will provide other services like curves, plotting, and summarization charts.

The publication manager (PM) is responsible to collect all the pieces of the publication together. It searches for the conference or journal paper template. The template will be filled initially with data from the Experiment database, for example the researchers' names, procedures and results. The generated plots and charts will be added to the template. At each step the Publication Manager will provide researchers with a preview for the paper. The researchers can stop, and save a draft or continue to the following steps. Publication manager allows groups of scientists to work on the publication at the same time from different locations, which promotes team science.

The citation manager (CM) looks up online databases (e.g. ACM or IEEE) and adds references to the reference section based on user selections; this will help solve some citation problems like author name ambiguity.

The feedback component will allow other researchers in the community to view the publications and write their

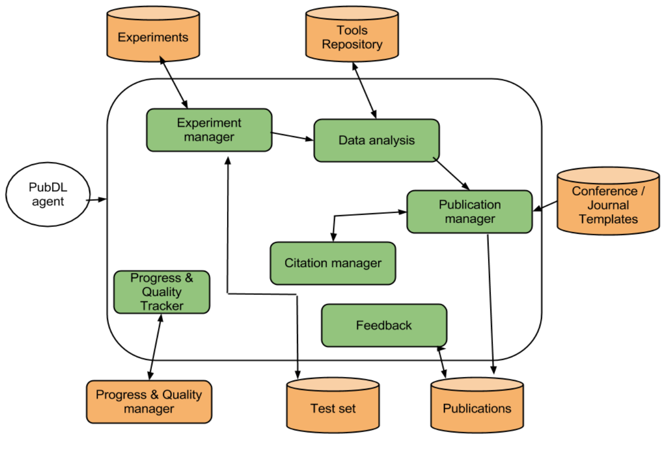

FIGURE 5
PUBLICATION DIGITAL LIBRARY

comments, which would help researchers to get early feedback before submission to conference or journal.

III. *Education Digital Library (EduDL)*

The EduDL is designed for educational and archiving purposes. The EduDL consists of four main components: The explorer, web crawler, virtual lab and course management system, as shown in figure 6. The explorer will provide access for two repositories: Bibliography (books, papers, journals, master and PhD theses, etc.), and Biographies (collections of useful web pages).

The user can search using keywords or browse using ontology. The course management system will host and maintain learning modules covering different topics in engineering education. Teachers can create, modify modules, and generate quizzes using question banks. Students will have access to course materials, and will able to engage different societies through discussion rooms. The virtual Lab is a collection of virtual reality tools, which simulate real labs. It will provide tutorials and scenario based learning activities.

IV. *Workflow Digital Library (WFDL)*

The WFDL is designed to store and maintain workflows created by users. The workflow [22] is to be described using formal or informal flow diagramming techniques, showing directed flows between processing steps. Single processing steps or components of a workflow can basically be defined by three parameters: input, output and processing. Using workflows will facilitate the researcher's work by hiding a lot of technical details and allowing the researcher to focus on the main concepts they are familiar with.

Researchers will use the Workflows to describe the experiments' steps. Users will be able to share their own workflows with others, combine two or more workflows into a new one, and add annotations and comments. Researchers can view the intermediate results between steps during the workflow execution or view the final results. The

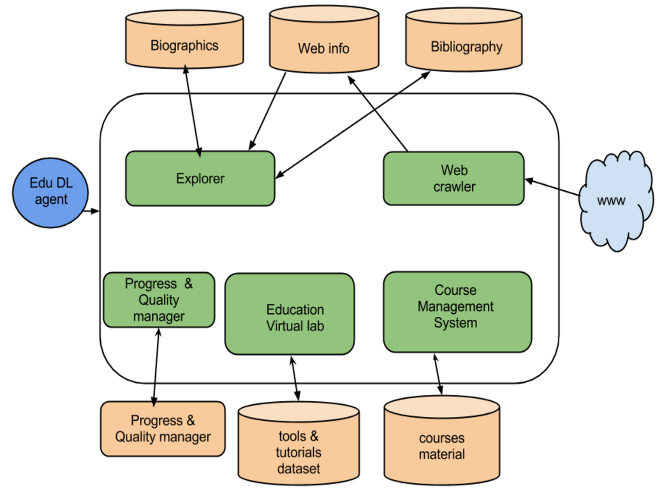

FIGURE 6
EDUCATION DIGITAL LIBRARY

Experiments results after workflow execution will be used later by the PubDL as part of the publication process.

V. *The Grid*

The Grid represents the core of TREE, where all repositories are stored and almost all of the data processing takes place. The Grid has many advantages:
1) No single point of failure.
2) Different copies of data on different machines.
3) Flexibility of upgrading and adding more storage and processing power.
4) Ease of resource accessibility.

The Grid consists of three main clusters dedicated to the digital library, as shown in figure 7. The organizer is responsible to process digital library requests. The request can be: execute job, check job status or transfer data. For the execute request, the digital library request should have some attributes like the target cluster or maximum time. Transfer data request can be to input or output data from or to the grid. The monitor checks periodically the status of each cluster, and provides this information to the organizer before each request. The cluster manager is responsible to submit the requests from the organizer to the workers. The workers are either computing worker (CW) or storage worker (SW). It is the responsibility of the cluster manager to put the plan for the workers to finish the assigned job.

THE 5S FORMALISM

There are several efforts to define standards for digital libraries, including the 5S and the DELOS frameworks. In this paper the 5S framework will be used to describe TREE. With 5S, our digital library including digital objects, metadata, collections and services can be easily described. It provides a specification based on five dimensions, including:

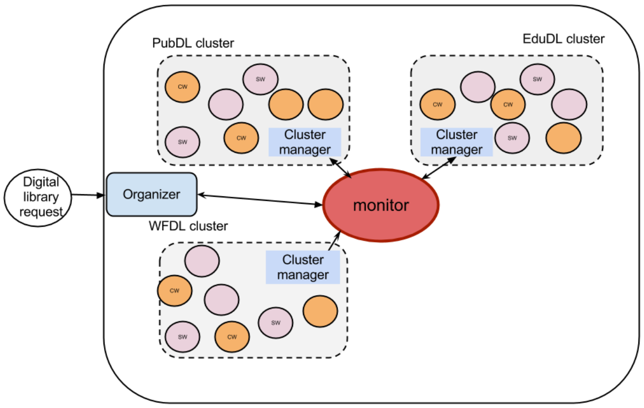

FIGURE 7
THE GRID

TABLE I
5S SUMMARY

5S	EduDL	WFDL	PubDL
Scenarios	Search Browse Archive Create course Create exam Exam	Create Workflow Execute workflow	Create experiment Plot/Analyze data Publish/ Cite
Societies	Student Teacher	Modeler Student Teacher	Researchers Teachers
Spatial	User Interface Index	User Interface Index	User Interface Index
Streams	Text Image Video	Workflow diagram	Text Image
Structural	Bibliography Course collection Document Hypertext Metadata	Workflow metadata Index	Conference/ Journal template Index Metadata Thesis Publication

1) Stream model, which describes properties of the DL content such as encoding and language for textual material or particular forms of multimedia data.
2) Structural model shows how data is organized
3) Scenarios model describes the behavior of library services and the way they interact together.
4) Societies model defines different types of users interacting with the library.
5) Spatial model provides the logical/presentation views of library components Table1 summarizes the 5S models in terms of their primitives, for the three libraries.

CONCLUSION AND FUTURE WORK

We proposed a conceptual design for TREE, a Digital Library for Engineering Education. TREE was designed as a collection of three digital libraries, each provideing different functionality and features. The Digital Libraries can be viewed as a network of Open Archives (OA). One of the main goals for our design was to hide all the technical details from users, where they can make use of the high performance computing capabilities provided by the grid which resides at the core of TREE. Other features include 24/7 availability, and access to wide range of online resources including courses, tutorials, multimedia collections, e-books and biographies. TREE will provide different services that promote research collaboration and team science amongst researchers by sharing their experiments' procedures and results with others. TREE will provide an online archive for research publications including journals, conference papers, dissertations and theses. As future work, the results of this work will be used to guide the implementation and testing of this system. Additional experiments that involve human subjects will be conducted to monitor, evaluate and assess the usage of this system by different users including students, faculty and researchers.

REFERENCES

[1] Gert, Janet. "*Selection for Preservation in the Digital Age.*" Library Resources & Technical Services. 44(2) (2000):97-104.

[2] Berners-Lee, Tim, Mark Fischetti, and Michael L. Foreword By-Dertouzos. *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. HarperInformation, 2000.

[3] Candela, Leonardo, et al. "*Setting the foundations of digital libraries*." D-Lib 13.3/4 (2007).

[4] Wright, Robert. "*The man who invented the web*." Time Magazine (1997): 64-68.

[5] Buckland, Michael K. "*Emanuel Goldberg, electronic document retrieval, and Vannevar Bush's Memex*." Journal of the American Society for Information Science 43.4 (1992): 284-294.

[6] Engelbart, Douglas C. "*NLS teleconferencing features: The journal, and shared-screen telephoning*." Fall COMPCON 75 Digest of Papers (1975): 173-177.

[7] E. A. Fox, Y. Chen, M. Akbar, C. A. Sha er, S. H. Edwards, P. Brusilovsky, D. D. Garcia, L. M. Delcambre, F. Decker, D. W. Archer, R. Furuta, F. Shipman, S. Carpenter, and L. Cassel. *Ensemble pdp-8: Eight principles for distributed portals*. In Proc. JCDL/ICADL 2010, June 21-25, Gold Coast, Australia, pages 341-344. ACM, 2010.

[8] Shaffer, Clifford A., et al. "*Building an online educational community for algorithm visualization*." Proceedings of the 41st ACM technical symposium on Computer science education. ACM, 2010.

[9] Dong, Andy, and Alice M. Agogino. "*Design principles for the information architecture of a SMET education digital library*." Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries. ACM, 2001.

[10] Women in Engineering ProActive Network , [Online]. Available: http://www.wepan.org/. [Accessed 5 June 2015].

[11] Northwestern Center of Engineering Education Research, [Online]. Available: http://nceer.northwestern.edu. [Accessed 5 June 2015].

[12] J. Weatherley, T. Sumner, M. Khoo, M. Wright, and M. Hoffmann, "*Partnership reviewing*," in Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries - JCDL '02, 2002, p. 106.

[13] Pandey, Richa. "*Digital library architecture*." DRTC Workshop on Digital Libraries: theory and practice. 2003.

[14] Carr, Leslie, et al. "*Extending the role of the digital library: computer support for creating articles*." Proceedings of the fifteenth ACM conference on Hypertext and hypermedia. ACM, 2004.

[15] Majithia, Shalil, et al. "*Triana: A graphical web service composition and execution toolkit.*" Web Services, 2004. Proceedings. IEEE International Conference on. IEEE, 2004.

[16] Ludäscher, Bertram, et al. "*Scientific workflow management and the Kepler system.*" Concurrency and Computation: Practice and Experience 18.10 (2006): 1039-1065.

[17] Oinn, Tom, et al. "*Taverna: lessons in creating a workflow environment for the life sciences*." Concurrency and Computation: Practice and Experience 18.10 (2006): 1067-1100.

[18] Peleg, Mor, Iwei Yeh, and Russ B. Altman. *"Modeling biological processes using Workflow and Petri Net models*." Bioinformatics (2001).

[19] Bartocci, Ezio, et al. "*BioWMS: a web-based Workflow Management System for bioinformatics*." BMC bioinformatics 8.Suppl 1 (2007): S2.

[20] Wu, Hulin, et al. "*Dediscover: a computation and simulation tool for hiv viral fitness research*." BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on. Vol. 1. IEEE, 2008.

[21] Lagoze, Carl, and Herbert Van de Sompel. "*The Open Archives Initiative: Building a low-barrier interoperability framework*." Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries. ACM, 2001.

[22] Deelman, Ewa, et al. "*Workflows and e-Science: An overview of workflow system features and capabilities*." Future Generation Computer Systems 25.5 (2009): 528-540.

AUTHOR INFORMATION

Sherif E. Abdelhamid PhD student at the Department of Computer Science, Virginia Polytechnic Institute and State University. Also affiliated with College of Computing and Information Technology, Arab Academy for Science and Technology, Egypt, sherief@vt.edu

Kenneth J. Reid Assistant Department Head for Undergraduate Programs and Associate Professor of Engineering Education, Virginia Polytechnic Institute and State University, kenreid@vt.edu