Volume 3, No. 11, November 2012
Journal of Global Research in Computer Science
ISSN-2229-371X
REVIEW ARTICAL
Available Online at www.jgrcs.info

WEB MINING AND THE EDUCATIONAL DOMAIN

Neha Gupta[*1] and Saba Hilal[2]

*[1]Department of Computer Applications, Manav Rachna International University, Faridabad, Haryana, India
Neha.fbc@mriu.edu.in[1]
[2]ISchool of Computer Sciences, Lingayas University, Faridabad, Haryana, India
Saba21hilal@gmail.com[2]

Abstract: The Web nowadays is a firmly established (virtual) reality that offers unprecedented opportunities to education. Many modes of delivery of online learning exist (e.g. educational software, virtual courses, blended learning, electronic books), all providing accessibility to learning materials, facilitating communication among learners and tutors/peers, and possibly helping to improve the learning and teaching process. While using an online learning environment, learners leave continuous hidden traces of their activity in the form of log file records, which document every action taken by three parameters: what was the action taken, who took it and when. The primary objective Web mining techniques is to use these log files, in order to portray the characteristics of the learner's learning processes in online environments illustrated by the portions given in this document. In this paper we have elaborated the use of web mining in various aspects of educational system.

INTRODUCTION

Web mining is a research area consisting of techniques and methods that automatically discover and extract information from Web log files [2]. Web mining has been massively used in e-commerce (e.g. in Amazon.com, where the customer's history is used for personalization and for encouraging further purchase), and is an emerging methodology in education research, assisting instructors and developers in improving learning environments and supporting decision-making of policymakers [1].

In education, the need for such methods appeared as early as the appearance of the Web. Unlike the use of Web mining in e-commerce - the target of which is to transform the surfer into a buyer - the same techniques in the context of e-learning aim to transform the learner into a more efficient learner [2]. In addition, usage analysis of e-learning is totally different from usage mining of e-commerce, since the learning process is far more complicated than the shopping process, and its cognitive aspects are much more difficult to track by means of log files.

WEB CONTENT MINING FOR EDUCATION AND LEARNING

Education is on the edge of a new epoch based on these changes. Online delivery of educational instruction provides the prospect to get colleges and universities new energy, students, and revenues. Many important educational institutions are working to start an online teaching and learning presence. Several special approaches have been developed to bring online education in an academic setting. For an extended era the use of web in education was limited because of the need for specialized hardware/software platforms for bringing that hypermedia to the end-user. For example the Plato system featuring personal and group notes, threaded discussions, hyperlinks, interactive elements and games, etc. Its use was restricted to institutes equipped with special terminals, connected to large mainframe computers running the Plato environment. A more reasonable but also ineffective attempt at bringing hypermedia and interactivity to the public was the introduction of CD-Interactive by Philips and Sony (in 1986). Many other (hypermedia) environments appropriate for delivering interactive learning material have been developed and never become popular.

Then, in 1989, Tim Berners Lee started developing the ideas and software for the World Wide Web. The first implementations of Web Servers (from CERN and NCSA, much later evolved into the most popular Apache server) and Browsers (NCSA Mosaic for X-Windows being the first popular one, later evolved into Netscape and now Mozilla Firefox) initially only allowed for a fairly primitive form of hypermedia: pages with (untyped) links, and possibly some embedded images. Soon a primitive form of servers-side dynamic content generation was added through CGI scripts (for Common Gateway Interface). Around 1993 people started publishing course material on the Web rather than in paper documents (or on proprietary systems in proprietary document formats) [7]. The Eindhoven University of Technology started the first on-line course on the subject of Hypermedia, using the Web as the hypermedia platform for delivering the course [7].

WEB BASED INTERACTIVE EDUCATIONAL HYPERMEDIA

Techno Interactivity is a central concept in educational environments [8]. It refers to the interaction of a learner with the learning material, the instructor, or with peers in the process of learning [9]. The basic question in relation to instructional design in the framework of e-learning and other computer-supported teaching and learning is how learners interact with educational multimedia; that is, what is their existing behavior, what their preferred learning style is, and what their learning goal in such an educational environment is.

In computer-based education, the interaction of a learner with what is to be learned, presented by the educational software, is central. Usually, educational software provides the learner with various forms of communication, often using different individual media. Therefore, we use the term interactive educational multimedia to emphasize the interactive nature of educational software and the variety of activities and interactions. Educational software can be seen as interactive

multimedia systems is a key to understanding learning and interactivity in computer-based educational environments. Multimedia software is designed for interaction with the user and with usability in mind. Principles of multimedia systems give details about the issues related to knowledge that is represented, activities that are offered, and types of interactions that are possible.

Multimedia systems are characterized by the communication channels that are provided for users to access and communicate knowledge. The user uses explicit languages to communicate along these channels. The interface that allows a (human) user to access and to communicate with the system plays a vital role. Channels and languages are central elements in the communication between agents in a multimedia system. An agent has an internal state, some goals and intentions, and the ability to communicate. Communication needs to be meaningful. We call a communication an interaction if it results in a change of state of the other agent. A common language that can be written or read by the agents is a prerequisite for interaction.

ANALYSIS OF WEB BASED EDUCATIONAL SYSTEM

Web-based education means that time and place are no longer barriers [10]. Any student who can get to a computer can take a web-based class and get an education at least equal to the one offered at a traditional MPA program [10]. In the United States and the other first world countries of the world, the limitation of computer access is not a major problem as they exist at the work place, in public libraries, and often in the homes of the students (Department of Commerce, 2001). The only serious computer problem that does exist is called "the last mile." This means that most computers are connected to the Internet using phone lines that have limited bandwidth, which results in a slow download time from the web. For the next decade, the solution for the professor is merely to post large files with video, such as Microsoft Producer files, to the web in two versions [10]. One posting is with high quality video. The other posting is with an audio only option. Then the professor lets the student decide which option is most suitable for them when calling up those files.

In others countries, computer barriers are more significant. For example, free access in libraries does not exist in many places and students are not as likely to have computers at home [10]. As in first world countries, bandwidth is also a serious handicap because the student will not be able to download large video files through telephone lines.

Several Online Education systems such as Blackboard, WebCT, Virtual University (VU), and some other similar systems have been developed to focus on course management issues. The objectives of these schemes are to present courses and educational programs through the web and other technologically enhanced media. These new technologies make it possible to offer instruction without the limitations of time and place found in traditional university programs. However, these systems tend to use existing materials and present them as a static package via the Internet.

Web based educational systems can be classified into two major categories:

1. Adaptive educational system
2. Intelligent educational system

Adaptive educational system: These systems attempt to be different for different students and groups of students by taking into account information accumulated in the individual or group student models.

Intelligent educational system: These systems apply techniques from the field of Artificial Intelligence (AI) to offer broader and improved support for the users of Web-based educational systems.

Both type of system are diverse but offer various kinds of support for both students and teachers involved in the process of Web-enhanced education. By adaptive and intelligent technologies we mean necessarily different ways to add adaptive or intelligent functionality to an educational system.

Based upon the above classification adaptive and intelligent Educational systems can be further classified as

1. Adaptive
- Adaptive hypermedia
- Adaptive Information Filtering

2. Intelligent
- Intelligent Monitoring
- Intelligent Collaborative Learning
- Intelligent Tutoring

The web based educational systems can further be analyzed on the basis of their origin and their representative systems as:

Table I. Web Based Educational Systems

Adaptive and intelligent educational systems	***Technologies***	***Sample systems***
Adaptive Hypermedia	• Adaptive navigation Support • Adaptive presentation	• AHA (De Bra, et al., 1998) InterBook (Brusilovsky, Eklund, & Schwarz, 1998) • KBS-Hyperbook (Henze, & Nejdl, • 2001) • MetaLinks (Murray, 2003) • ActiveMath (Melis, et al., 2001) • ELM-ART (Weber, & Brusilovsky, • 2001) • INSPIRE (Papanikolaou, Grigoriadou, • Kornilakis, & Magoulas, Submitted)
Adaptive Information Filtering	• Content-based filtering • Collaborative filtering	• MLTutor (Smith, & landford, 2003) • WebCOBALT (Mitsuhara, et al., 2002)

Intelligent Class Monitoring		• HyperClassroom (Oda, Satoh, & Watanabe, 1998)
Intelligent Collaborative Learning	• Adaptive group formation and peer help • Adaptive collaboration support (coaches and monitors) • Virtual students	• PhelpS (Greer, et al., 1998) • HabiPro (Vizcaíno, Contreras, Favela, & Prieto, 2000) • COLER (Constantino Gonzalez, • Suthers, & Escamilla De Los Santos, 2003) • EPSILON (Soller, & Lesgold, 2003)
Intelligent Tutoring	• Curriculum sequencing • Intelligent solution analysis • Problem solving support	• VC-Prolog-Tutor (Peylo, Teiken, • Rollinger, & Gust, 1999) • SQL-Tutor (Mitrovic, 2003) • German Tutor (Heift, et al., 2001) • ActiveMath (Melis, et al., 2001) • ELM-ART (Weber, et al., 2001)

WEB CONTENT MINING AND DISTANCE LEARNING

Distance education or distance learning consists of techniques and methods providing access to educational programs for students who are divided by time and space from lecturers. E-Learning systems do not have a closer student–educator relationship (one to one). There are different subtypes of distance education: paper-based correspondence education, videotape education, computer aided education (multimedia education, internet education or web-based education), etc. Currently, the most used is web-based education allowing students to conveniently learn via the Internet. Web-based education is a form of distance education delivered over the Internet. Today, there are a lot of terms used to refer to web-based education such as e-learning, e-training, online instruction, web-based learning, web-based training, web-based instruction, etc. And there are different types of web-based systems:

1. Synchronous and asynchronous
2. Collaborative and non-collaborative
3. Closed corpus and open corpus, etc.

These web-based education systems can normally record the student's accesses in web logs that provide a raw trace of the learner's navigation on the site.

There are various types of web mining techniques applied to educational systems, but nearly all of them can be grouped in one of the five next ones:

1. Clustering
2. Classification
3. Outlier detection
4. Association rule mining and sequential pattern mining
5. Text mining.

The need to personalize distance education courses stems from their ultimate goal: the need to serve an individual student independently of time, place, or any other restrictions [11]. This means that a distance education system must be served by a web content mining systems to monitor, intervene in, and counsel the teaching, studying- learning process. Compared to current intelligent tutoring systems or adaptive learning environments where a teacher has only an occasional role, web content mining emphasizes the role of expert, in this case teacher, to interpret the findings obtained from analyzing the data retrieved from the course [11].

Web mining can help perform the following task to improve the distance education on the web:

1. To understand student behavior,
2. To determine e-learning system effectiveness,
3. To measure the success of instructional efforts.

Organization that is responsible for distance education collect huge volume of data, which are generated automatically by web servers and collected in the server access logs [12]. They also gather information from referrer logs which include information about the referring pages for each page and also from user registration. Through this an organization can get idea about thinking styles, learns their expectations and also about the web site structure. This helps to improve the efficiency of the web site that is responsible for improving the knowledge of the learners.

Web mining in distance education provides a lot of open teaching resources, so that people can teach and learn anytime and anywhere [12]. It helps the organization that is responsible for distance education to discover the learner‘s access habit and the study interest. It directs the teacher to change his/her teaching techniques and the speed of teaching depending on the learner‘s knowledge. So web mining technology is a key enabler of distance education.

VIRTUAL WEB VIEW [13]

The VWV plays the role of a data warehouse for web content. A VWV provides a window to observe a subset of Web resources, and gives the illusion of a structured world. It is an approach is to construct progressively a global multiple layered database by generalization and transformation of lower layered data, store and manage multiple layered information by database technology, and perform resource and knowledge discovery by query transformation, query processing and data mining techniques. A VWV is based on concept hierarchies and it is very difficult to find a consensus on a general ontology. It is more realistic to build different VWVs specializing in different topics or restricted geographically, etc. VWVs can also share the same primitive data but use different ontology's (i.e. concept hierarchies). In such a context a software agent that plays the role of a mediator and broker between VWVs is necessary.

The major strength of the VWV approach is its promotion of a tight integration of database and data mining technologies with resource and knowledge discovery in global information systems. With the dynamically growing, highly unstructured, globally distributed and huge information base, the application of the mature database technology and promising data mining techniques could be an important direction to enhance the power and performance of global information systems.

Virtual web view concept

1. Multiple layered database (MLDB) is used to handle large amounts of unstructured data on the Web.
2. This database is massive and distributed. Each layer is more generalized than the layer beneath it.
3. The MLDB provides an abstracted and condensed view of a portion of the Web. A view of the MLDB, which is called a Virtual Web View (VWV), can be constructed.
4. Generalization tools are proposed, and concept hierarchies are used to assist in the generalization process for constructing the higher levels of the MLDB.
5. WebML, a web data mining query language is proposed to provide data mining operations on the MLDB. It is an extension of DMQL**.**

VIRTUAL EDUCATION AND WEB MINING [6]

Virtual education is in an attempt to create an alternative to traditional education environments. Virtual education environments depict a rapid development together with the developments in the internet technologies. While virtual education environments proceeded in a one way structure during 1990s, it is seen that virtual education environments have presented bidirectional interaction and completely 3 dimensional study environments by 2000s. It is also seen that virtual education environments are in an attempt to create an alternative to traditional class environments by virtual libraries, virtual quizzes, guidance services, smart content systems and education management systems. However, it is seen that among all these elements, the greatest lack is the knowledge with no meaning.

Virtual education can be considered as of three categories consisted of educational activities in university level for now.

These are:

1. In-service education programs for occupation purposes,
2. Academicals degree programs
   a. With undergraduate programs
   b. Graduate (post graduate and doctorate) programs
3. Special programs for social purposes [3]

Together with the developments of internet technologies, more online communication and education environments have begun to be used. The conveniences brought particularly by the new technologies presented freely and the tools of web 2.0 which could be considered as an internet revolution have begun to be used and as parallel to this, the concept of Learning 2.0 has emerged [4]. Virtual education environments produced for educational purposes are commonly in different system cases of Virtual Learning Environments (VLE), Learning Management Systems (LMS), Course Management Systems (CMS) and Learning Content Management Systems (LCM) [5]. Nowadays, education environments have two forms. These are traditional and virtual education environments. In both environments, one of the important problems is habit analysis and evaluation. In traditional education systems, habit analysis is done easily through observation techniques, whereas in virtual ones analyzing students' attitudes and habits is a significant problem. Web mining applications, used in providing meaningful information from meaningless habits and surfing in web environments, is the way to overcome this problem.

REFERENCES

[1] Cohen, A., & Nachmias, R. (2006). A quantitative cost effectiveness model for Web-supported academic instruction. The Internet and Higher Education, 9(2), 81-90. (2006).

[2] Rafaeli, S., & Ravid, G. (1997). Online, Web-based learning environment for an Information systems course: Access logs, linearity and performance. Paper presented at the Information Systems Education Conference, Orlando, FL.

[3] Karasar, (2004). Eitimde Yeni letiim Teknolojileri, The Turkish Online Journal of Educational Technology - TOJET October 2004 ISSN:1303-6521 Volume 3, Issue 4, Article 16

[4] Karrer, T. (2007). Understanding eLearning 2.0. Retrieved 12 05, 2009, from http://www.astd.org/LC/2007/0707_karrer.htm

[5] González, M., Hernández, G., & Martínez, J. H.-R. (2009). Learning objects in theory and practice: A vision from Mexican University teachers. Computers & Education, 1330.

[6] Tuncay Sevindika *, Necmi Demirkeserb, Zafer Cömertb, "Virtual education environments and web mining" , WCES-2010, Procedia Social and Behavioral Sciences 2 (2010) 5120–5124

[7] P. De Bra, "Web-based educational hypermedia", Communications of the ACM, Volume 45 Issue 5, May 2002 Pages 60 – 61

[8] Sims, R., "Interactive learning as "emerging" technology: A reassessment of interactive and instructional design strategies" , Australian Journal of Educational Technology, Vol 13(1), Pg 68-84, (1997).

[9] Moore, M.G, "Three types of interaction", The American Journal of Distance Education, Vol 3(2), Pg1-6, 1992.

[10] Thomas D. Lynch and Cynthia E. Lynch, "Web-Based Education", The innovation Journal, Vol 8(2) (07/10/2003)

[11] W. H¨am¨al¨ainen, T. H. Laine, E. Sutinen, "Data Mining in Personalizing Distance Education Courses", Department of Computer Science, University of Joensuu, WIT Press, Hong Kong (pp. 1–11)

[12] D.Santhi Jeslet Dr . K.Thangadurai, "Web Mining: A Key enabler for Distance Education", GJCST, Vol. 10 Issue 5 Ver. 1.0 July 2010

[13] Osmar R Ziane, "Warehousing the Web or Building Virtual Web Views", In Proc. Conf. on

SHORT BIODATA OF ALL THE AUTHOR

Ms. Neha Gupta is presently working with Manav Rachna International University, Faridabad as Assistant Professor and is Pursuing PhD from MRIU. She has completed MCA from MDU, 2006 and Phil from CDLU, 2009. She has 6 years of work experience in education and industry. She is faculty coordinator and has received many appreciation letters for excellent teaching. She has published 8 papers in many reputed international journals which are indexed by scopus.