Proceedings of the 10th Panhellenic Conference in Informatics, PCI2005, 11-13 November, Volos, Greece, pp. 270-279, 2005.

On the Personalization of an E-learning Environment in Higher Education using Graphs

L. Drossos[1], K. Giotopoulos[3], B. Vassiliadis[2], E. Sakkopoulos[3], K. Votis[3], A. Stefani[2], A. Tsakalidis[3], M. Xenos[2], S. Likothanassis[3]
[1]Technical Educational Institute of Messologi
{ldrossos@teimes.gr}
[2]School of Science and Technology, Hellenic Open University
{bb,stefani,xenos}@eap.gr
[3]Dept. of Computer Engineering and Informatics, University of Patras
{kgiotop,sakkopul,botis}@ceid.upatras.gr

Abstract: Blended or enhanced learning models refer to the parallel use of traditional and on-line teaching methods. Recently, web based systems are gaining respect as enablers of such models due to their cost-effectiveness. In this work we present some preliminary results from a study of workspace system use for blended learning in a higher education institute. Our experience with the introduction of this service is based on access log analysis in the frame of a computer science undergraduate course. The evaluation sought to measure the success of this pilot application with several criteria. Other technological and organization issues involved are also discussed along with possible extensions that lead to a fully personalised knowledge workspace.

1 Introduction

User driven access to information and services has become more complicated, and can sometimes be tedious for users with different goals, interests, levels of expertise, abilities and preferences [1]. The process of designing, developing and maintaining e-learning applications based on web technologies is also a challenging task. Besides technical difficulties, parameters such as learner diversity, continuous change and increasing student population pose stressing requirements for automatic and cost-effective adaptation. Usage pattern recognition is one of the most popular and, at the same time, difficult techniques used for adaptation of web applications. Especially in web-based learning, mining of usage patterns is usually biased by irrelevant or misleading trails. There are numerous approaches to e-learning but very few are efficient when applied to real world situations [21].

Usage pattern recognition, as a web adaptation enabler, is driving the change from content-centric to user-centric e-learning systems. Although true personalisation is not yet in our grasp, especially in large e-learning corpuses, significant advances have already been presented [2]. In this work we focus on a significant parameter that affects usability: orientation and navigation strategy. Users are frequently uncertain as how to reach their goals. Since users have different states of knowledge and experi-

ence, information presentation may be too redundant for some and too detailed for others.

In this paper we present an approach for the implementation of the enhancement of an existing off-line collaboration system for blended learning with personalisation/adaptation capabilities. Intelligent, adaptive and personalised services are proposed through access log analysis and discovery of user trails and patterns. The goal of our work is to discuss the proposed environment and the techniques used to allow personalization and adaptation, and how such features could be used to reach a simple, yet powerful and cost-effective web assistance environment.

The structure of the paper is as follows: in section 2 an overview of related techniques and applications is provided. In section 3 the existing application is described and the new enhancements are discussed. Section 4 analyses the personalisation strategy of the system: algorithms for log file analysis are described in detail. Finally, section 5 discusses future directions.

2. Adaptation in web based e-learning

Web applications rely on a “pull” model of information flow and thus users are expected to drive the adoption of new technologies. Although, adaptable systems have been considered in a wide range of recent research efforts, they are mainly focused on e-commerce. E-learning adaptation has many common characteristics with e-commerce applications and many differences as well. This means that advances in e-commerce adaptation may only partially used for e-learning adaptation [3,4]. In this section we initially overview recent advanced in web-systems adaptation and afterwards we examine specific techniques for web based e-learning log analysis.

Most existing adaptation techniques for web-based systems are based on log analysis, user modelling and/or pre-determined rules. Rules assign adaptation constituents to interaction situations in a rather "arbitrary" way. These rules are usually hard-coded in the user interface, and cannot be easily modified, or reused across different applications. Additionally, there are no explicit representation of the goals underlying adaptation processes, and, in this sense, the latter cannot be taken explicitly into account in the adaptation process. Decision-theoretic frameworks for run-time adaptation, are mostly utility-based decision making techniques in the context of the standard reference model for intelligent hypermedia presentation systems [5].

Static user profiles are the primary source of knowledge in most recent implementations of personalized information retrieval systems and adaptive e-shop applications. Profiling is based mostly on explicit graded user feedback and in some occasions on server logs analysis [6]. Although this approach may be suited for constraint space applications or situations where detailed user information is available, it is not efficient for supporting first time users. Implementations that simulate human assistants provide help to users that browse through the catalogues of e-shops, by taking advantage of multi-agent architectures and detailed user profiles [7]. Researchers are also looking at the potential of Bayesian networks and genetic algorithms as mechanisms for agents to learn more complex and effective adaptation.

Recently, web engineering approaches [9] have been used in order to reduce cognitive overload in hypermedia environments based on navigational patterns, semantics and social interaction [8] have been proposed.

Web log analysis and mining of user patterns is a hot subject for e-commerce applications. Research efforts have already bear fruits: real life e-commerce systems, such as the recommender engine of Amazon.com, are successfully providing adaptation in the form of recommendations. Marketing systems are using machine learning techniques to analyse email text and discover user preferences. Business intelligence, an information systems concepts, is expanded to include web based systems. A plethora of log analysis techniques has also appeared in the literature showing good adaptation results. One might expect that e-learning could use directly the same (or slightly modified) methods for adaptation. In the case of adaptation using log analysis (the focus of this work) this has not happened yet in a large scale. Web based e-commerce and web based e-learning may seem analogous but have very significant differences. For example, a session in e-learning is very different from a transaction (session) in e-commerce. In the latter case, the session begins with an access to a web page and ends with an order, a logout from the electronic basket. In any case, their duration is short. An e-learning session may span many access sessions (e.g. access forums, download exercises, take online tests etc.). Collaboration, and thus links between users, is stronger in e-learning. For example forums are more active in e-learning than in an e-shop. Goals are also different in many aspects. E-commerce goals are more clear: keep the customer satisfied and encourage him to buy more products, although in a discrete way [10]. E-learning specific goals are more difficult to pinpoint: they are more subjective and strongly context dependent. Furthermore, new trends in pedagogy concentrate more on constructivism, the building of knowledge by way of social interaction and collaboration on-line [11]. Although constructivism has been identified by many researcher as one of the most appropriate methods for learning, science has not yet comprehended and analysed the mental processes of human knowledge building, sharing, evolution and reuse. Thus, a learner does not behave exactly as a buyer does.

To receive web usage feedback, e-learning web sites have been accompanied with logging mechanisms that have been evolving over time. However, the consequences of ad hoc implementation are depicted on the logged navigation trails, where mining for useful information in the logs has become a travel through an information labyrinth. A shift from standard HTML based applications toward server side programmed web applications is noted several years now, especially with the advent of technologies such as Java servlets, PHP and lately with Microsoft .NET. Among other features, new techniques allow URL re-writing to provide additional information on the HTTP requests, HTML server-side pre-rendering or pre-compilation to facilitate quicker download, client-side code injection to enable transparent capturing of additional user actions and custom logging databases that keep details regarding content delivery to web and proxy servers.

Significant work on converting server logfiles to valuable sources of access patterns has been conducted by Cooley [12]. Apart from analysing logfiles, it is important to use analysis as input and determine which changes, if any, to bring to the website structure. Srikant and Yang [13] infer path traversal patterns and use them to

indicate structural changes that maximize (or minimize) certain site-dependent criteria. Finally, in [14], [15] the authors define techniques to assess the actual value of web-pages and experiment on techniques and mechanisms to reorganize websites. Personalized customary e-learning solutions include [16], [17], [18] which have been developed for the University of Patras and the Advanced School of Fine Arts respectively.

3. A web assistance system for higher education

The Technical Educational Institute of Messologi is one of the leading Higher Technical Education institutions in Greece with a student population of more than 4000 students, supporting 16 graduate courses in 4 different departments. The Department of Applied Informatics in Management & Economy has recently decided to deploy a set of on-line collaborative tools enabling an enhanced teaching model: traditional teaching methods and on-line support.

In this endeavour, development, support and expansion costs were important considerations. In this work we argue that standard web technology, if properly used, can provide a cost-effective means for enhanced learning even in higher education environments. It must be noted that on-line tools are only used to support, not replace traditional methods.

The general architecture uses a portal as a single access point to any current or future tools or services that support the new blended educational model. The heart of the system, in its current state, is the web based course assistance tool and an application based on the MOODLE on-line cooperative tool. User services include basic information sharing and management – (create and publish the latest course material for immediate use), searching (navigation – finding information from any location within or from outside the network), communication and collaboration (mainly using off-line web tools such as forums). For example, figure 1 presents forum announcements.

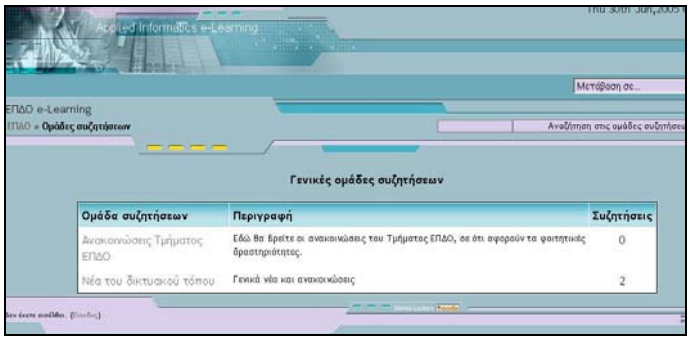

Figure 1: Applied Informatics e-Learning Announcements

A very significant characteristic of the system is that a part of the on-line information is grouped based on context: courses and laboratories. This includes both course material and off-line collaboration content (messages, announcements etc.). The use of forums and course-related information creates two user communities: the online communities and the course communities. Although they both share many characteristics, course communities are more formal while on-line communities do not have strong ties between their members: they are formed and dispatched more easily. Virtual community support will be enabled in the future by the creation of individual student workspaces using appropriate tools (e.g. BSCW). Although we do not expect that on-line virtual community tools will have significant success, it would be interesting to evaluate their use in such a blended environment. Figure 2 presents the basic form for enlisting to a course.

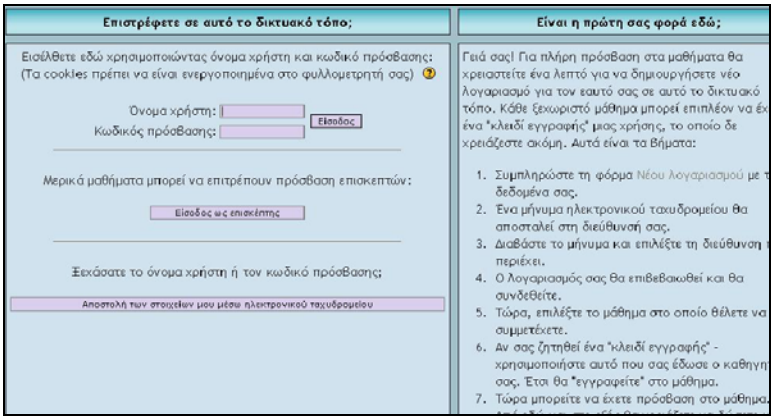

Figure 2: Form for course enlistment

The adaptation module (not available at this time) will be used for presenting the user with an HTML document composed of three parts:

- Adaptive node presentation: enables the information content of hypertext nodes to adapt to the user's expectations or requirements,
- Adaptive node navigation: recommendation of links likely to lead to information relevant to the primary task of the user, or by changing the destinations of links,
- Main page: This part of the page remains constant for all users.

It must be noted that only a portion of each page will be adapted to the user preferences. This depends on available user information. From the two adaptable parts of each page described above, the most demanding, as far as user information is concerned, is the Adaptive Node presentation part. An adaptation mechanism is responsible for dispatching orders to the user interface (residing at the portal). Adaptation orders are formed based on log analysis processing results that mine information such as:

- the most popular page for this user
- the most popular page for the user group the user belongs to
- the most popular page/message/file per course the student is registered to

Apart form individual user information, there is an additional source that is of important use to this system: the preferences of the user communities. Users can be grouped into communities based on their system asset preferences (strategy preferences, upload/downloads, registered courses, page hits etc). Recommendations of communities may be weighted accordingly and taken into account, especially in the decision making process where individual user data are not adequate for personalization. The log analysis technique that will be used for the realization of this mechanism is discussed in more detail in the following section.

4. Usage pattern recognition using graphs and some preliminary results

4.1 Visualization software

This section describes functional characteristics of the visualization software utilized during the platform evaluation. Extensive details are presented in [19].

The software is able to perform structural and statistical analysis of a given website, using a GUI (see Figure 3). It supports a number of log input types such as W3C, IIS, NCSA, ODBC logs in typical or extended format. An HTML parser and a link crawler are included, which once receiving the homepage of a website discover its link structure. Furthermore, the user can provide, in the corresponding window, the website logfile. The logfile is analysed to discover "Maximal Forward Paths" and "Forward Paths", as defined in [20]. The user can choose to display the site graph and the most frequent paths. In this way, possible recommendations can be identified quickly and directly on the site visual representation.

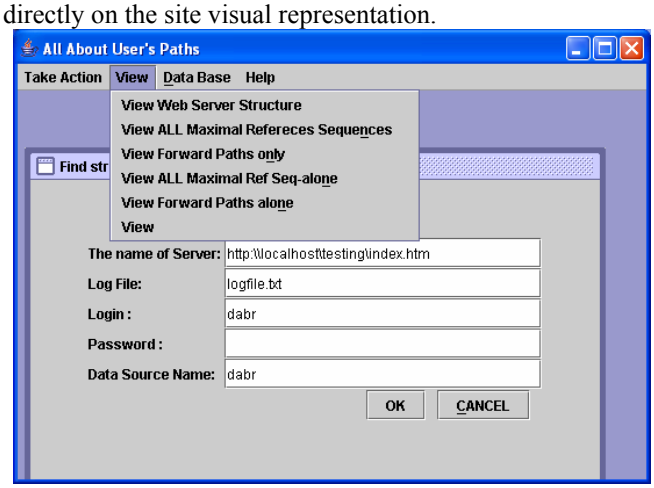

Figure 3: Visualization User Interface [19]

4.2 Using routing probabilities

In our analysis, we have utilized the routing probabilities technique to receive the relative access count from the logs as introduced in [14]. In [14] it is observed that a web page is accessed in four different ways. First it gets accesses within site, second directly via bookmarks, thirdly by incoming links from the outside world and finally by typing directly, its URL.

In routing probabilities as presented in [14], the idea was to increase page relative weight, inversely proportional to its access probability. Let's think of the site structure as a *directed acyclic graph* (DAG) G, with s nodes and v_i denoting page i. Suppose that a user starts at the root page v_r, looking for an arbitrary site page v_t. At each node v_i he makes two kinds of decisions: either he stops browsing or he follows one of the $out(v_i)$ links to pages on the same site. If we consider each kind equally probable, the probability p_i, of each decision is $p_i = (out(v_i)+1)^{-1}$.

Consider a path $W_j = \{v_r, v_1, \ldots, v_t\}$, from v_r to v_t. Counting the routing probabilities at each step, the probability of ending up to v_t via W_j, is:

$$P_{t,j} = \prod_{\forall i, v_i \in W} p_i$$

There may be more than one paths leading to t, namely $W_1, W_2, \ldots, W_k$. The overall probability of discovering t, D_t is:

$$D_t = \sum_{i=1}^{k} W_i$$

Considering page i as target, the highest D_i is the lower $a_{i,in}$ shall be, so we choose $a_{i,in}$ to be, $a_{i,in} = 1 - D_i$. We also let $a_{i,out} = 1$. Thus we define RA_i as:

$$RA_i = (1 - D_i) \cdot AA'_{i,in} + AA_{i,out}$$

with $AA'_{i,in}$ and $AA_{i,out}$ defined as the incoming and outgoing absolute number of accesses logged.

4.3 Some preliminary experimental results

The educational system provides e-Learning services for 643 distinct students and faculty registered in all categories. There are sixty nine (69) courses/laboratories introduced in the platform that facilitate the autumn/winter semester in the academic year 2004-2005. Fifty eight (58) of them have shown significant traffic beyond the average, while the rest electronic courses have been only partially adopted during the educational procedure. Overall, weekly log analysis has shown that the system successfully delivers e-Learning content for at least 99,74% of the incoming hits, which indicates that the system overall has a stable network connection and satisfies the users' needs while browsing.

Average visits per course are 230 on weekly basis. However, the categorization of the users' action is depicted in figure 4. As expected, users have mainly utilized the course related material published by the e-Learning system directly and they have shown only limited interest in our features of the system such as discussion/announces, searching and account personalization & administration.

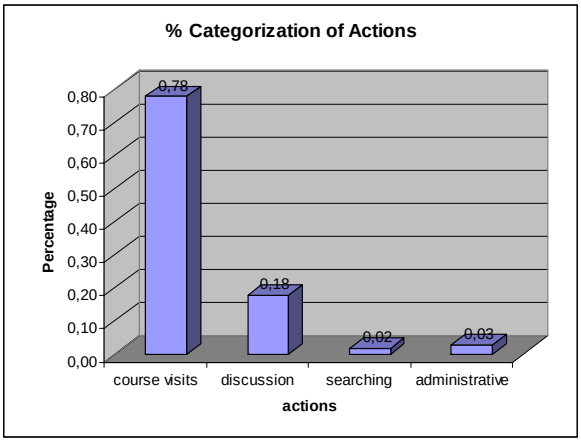

Figure 4: % Categorization of Actions: Course/ Lab visits are dominating

Further investigation and analysis of the e-Learning system's log will be presented in the future including even semester courses/ laboratories activities.

5. Conclusions

Web based technology is the technology of choice for e-learning due to its cost-effectiveness, its simplicity and its flexibility. New blended or enhanced models use traditional teaching methods combined with static or dynamic tools based on simple web technologies.

In this work we presented the first steps of such an endeavour in a higher education institution. In summary, we have discussed the use of log files and log analysis as a starting point in developing adaptive assisting tools for blended learning. Our approach will use Maximal Forward Paths derived from student log analysis in order to adapt the user interface: to recommend popular links, files and actions.

The ultimate aim of our work is to explore how we can fully integrate adaptation and tutoring techniques in a computer-mediated collaborative environment. In other words, to use the integration of personal workspace and low-cost off-line collaboration tools as a first step toward developing a fully integrated, low cost environment.

Acknowledgements

This work was supported by the Greek Secretariat for Research and Technology Iraklitos programme (contract no: B.238.027).

References

1. Allison, C., Cerri, S.A., Ritrovato, P., Gaeta, M., Salerno, S.: Human Learning as a Global Challenge: European Learning Grid Infrastructure. University of Tampere Press (2003)
2. Hiltz, S.R., Turoff, M.: What makes learning networks effective?. Communication of the ACM Vol.45(4) (2002) 56-58
3. Schwier, R.A.:Virtual learning communities. In G. Anglin (Ed.), Critical issues in instructional technology. Portsmouth, NH: Teacher Ideas Press (2004)
4. Jonassen, D.H., Howland, J., Moore, J., Marra, R.M.: Learning to solve problems with Technology. Pearson Education (2003)
5. Grønbæk, K. & Trigg, R.H.: From Web to: Designing Open Hypermedia Systems. MIT Press ISBN 0-262-07191-6 (1999) 424
6. Gams, E., Reich, S.: An analysis of the applicability of user trails in web applications. Web Engineering Workshop (2004)
7. Kim, J., Oard, D. W., and Romanik, K.: Using implicit feedback for user modelling in Internet and Intranet searching. Technical Report, College of Library and Information Services University of Maryland at College Park (2000)
8. Dieberger, A.: Social Navigation in Populated Information Spaces. In A. Munro, K. Höök & D. Benyon (eds.): Social Navigation of Information Space London: Springer (1999) 35-54
9. Murugesan, S., Deshpande, Y.: Meeting the challenges of web application development: the web engineering approach. Proceedings of the 24th international conference on Software engineering ACM Press (2002) 687–688
10. Li, L. Zaiane, O.R. Combining Usage, Content and Structure Data to Improve Web Site Recommendation, 5th International Conference on Electronic Commerce and Web Technologies (EC-Web 04), Springer Verlag LNCS 3182 (2004) 305-315
11. Bumpous, D.: Constructivism and E-learning Applications. ELRN 772 (2003)
12. Cooley, R.: Web Usage Mining: Discovery and Application of Interesting Patterns from Web data. PhD thesis University of Minnesota (2000)
13. Srikant, R., Yang, Y.: Mining Web Logs to Improve Web Site Organization. Proc. WWW01 (2001) 430-437
14. Christopoulou, E., Garofalakis, J., Makris, Ch., Panagis, Y., Sakkopoulos, E., Tsakalidis, A.: Techniques and Metrics for Improving Website Structure. Journal of Web Engineering, Rintton Press, JWE030416, Vol.2 (1&2) (2003) 90-104
15. Christopoulou, E., Garofalakis, J., Makris, Ch., Panagis, Y., Sakkopoulos, E., Tsakalidis, A.: Automating Restructuring of Web Applications. ACM Hypertext 2002 (2002)
16. Bouras, C., Destounis, P., Garofalakis, J., Gkamas, A., Sakalis, G., Sakkopoulos, E., Tsaknakis, J., Tsiatsos, Th.: Efficient Web-based open distance learning. Telematics and Informatics Article No. 171 (2000)
17. Bouras, C., Destounis, P., Garofalakis, J., Gkamas, A., Sakalis, G., Sakkopoulos, E., Tsaknakis, J., Tsiatsos, Th.: Efficient Data Management to support Educational Activities. International EurODL: European Journal of Open and Distance Learning, ISSN# 1027-5207 (1999)

18. Sakkopoulos, E., Tsakalidis, A.: Utilizing Complementary Know-How: Advanced Fine Arts meeting Information Technology to provide a Virtual University for artists - students and alumni. ACM SIGUCCS User Services conference (2002)
19. Avramouli, D., Garofalakis, J., Kavvadias, D.J., Makris, C., Panagis, Y., Sakkopoulos, E.: Popular Web Hot Spots Identification and Visualization. Fourteenth International World Wide Web Conference 2005 (WWW2005) (2005)
20. Chen, M.-S., Park, J.S., Yu, P.S.: Efficient Data mining for path traversal patterns. Knowledge and Data Eng. Vol.10(2) (1998) 209–221
21. Xenos M., Pierrakeas C. and Pintelas P.: Survey on Student Dropout Rates and Dropout Causes Concerning the Students in the Course of Informatics of the Hellenic Open University. Computers & Education Vol. 39 (4) (2002) 361-377