

SUPPORTING DECISION OF MANAGEMENT MAKING VIA WEB MINING IN EDUCATIONAL INSTITUTIONS

László Bóta

Eszterházy Károly College, Eger, Hungary
E-mail: botal@ektf.hu

Abstract

While the management of education institutions is interested in gaining a greater level of knowledge of learner habits and behaviour, in case of students using on-line communication devices this goal cannot be achieved with traditional means.

Management aspirations for the provision of services satisfying the information demands of students and visitors can only be realised if the users' habits and behaviour are analysed according to the respective links reached. The visitor link request registers contain adequate amount of relevant and valuable data for the development of ergonomic quality and the personalization of the on-line communication surface. Web mining can help in retrieving this information.

The principal subject of the present examination is the home page of the Eszterházy Károly College. The aim of the research effort is the enhancement of the software-ergonomic quality of this home page. During the examination of the web journal entries we try to identify the accessibility patterns of the respective web pages in order to improve the quality of Internet-based services.

Key words: *on-line communication, software ergonomics, student behaviour and habits, web mining.*

Introduction

The home page of the Eszterházy Károly College has been in operation for 12 years and is visited by approximately 5 to 8,000 people a week. The visitors are mostly students, instructors, and those interested in continuing their studies at the institution. However, in lieu of a thorough analysis the management of the institution has not become fully aware of the problems and habits of on-line visitors.

Management decisions based on software ergonomic examinations and the resulting developments facilitate continuous quality assurance. While the home page of the Eszterházy Károly College has undergone numerous changes and repeatedly has been expanded throughout the years, a comprehensive software-ergonomic study concerning the users and the efficiency of the respective documents located on the webpage has not been carried out yet. This essay will introduce the beginning phase of this comprehensive analysis.

Since the establishment of the College's web-page on June 26, 1996 its profile changed significantly three times (1998, 2002, 2007). The last profile change took place as of 10.00 AM October 9, 2007. While this time the fundamental errors identified by the analysis of the previous home page were avoided, identification of the imminent potential short comings appears to be an important task as well.

The subject of the examination efforts

The first quantifiable results are derived from a comprehensive and simultaneous examination of the current and previous home page. The use of exclusively web mining-oriented technology aimed at the identification of potential sources of misinformation.

The first inquiry

The first examination focuses on the *Instruction* menu point of the left side of the old home page of the Eszterházy Károly College. While a click on the menu should reveal the training profile, the respective major's programs, departments, course units, and syllabi, only information pertaining to a smaller group of visitors concerning complementary programs and special courses is accessible. The exact time frame of the examination focuses on January-February 2007.

We attempt to prove the following hypothesis: The *Instruction* menu point is misleading as due to the deceptive denomination the visitor can be misrouted.

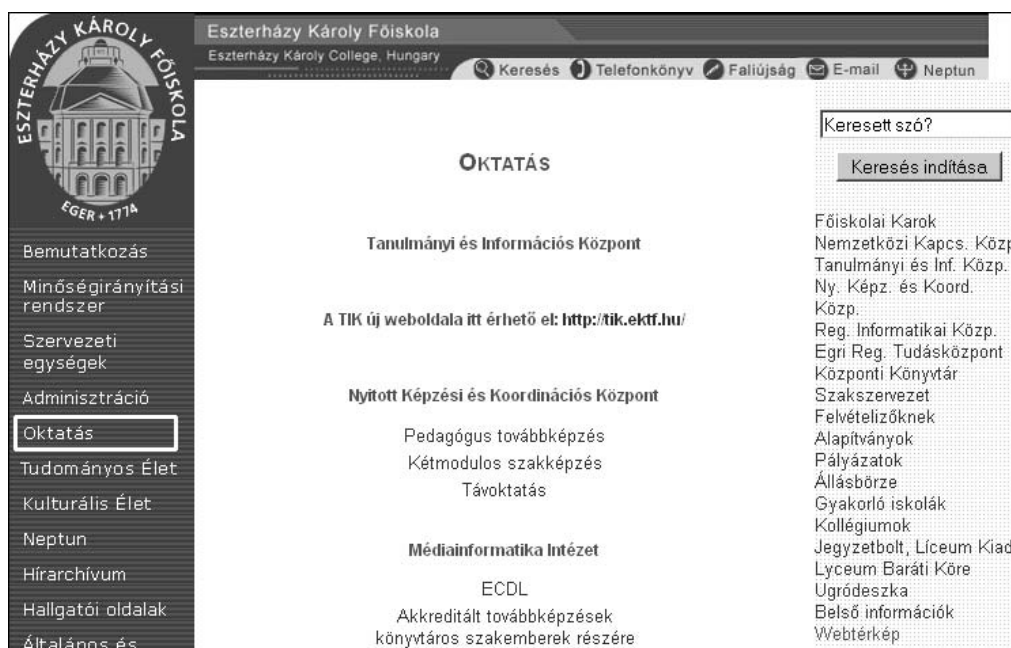


Figure 1. The lay-out of the web surface after clicking on the *Instruction* menu point located on the old home page.

The second inquiry

In addition to the elimination of the potentially misleading capacity of the *Instruction* menu point we aimed at the identification of other shortcomings. While the new home page is thoroughly designed, the respective numbers reveal a potential user deception capability. In order to perform a full examination the correlation among all menu points would have to be analyzed. This however, due to our familiarity with the content of the web page and the respective topics is not necessary. The second examination is based on the menu structure of the new home page (Figure 2). The examination focusing on the period of October 9-24, 2007 aims at the identification of user disorientation problems in the menu structure.

Figure 2. The lay-out of the new homepage of the Eszterházy Károly College.

Web mining as an examination method

The World Wide Web and the relevant divided information providing services function as an educational surface as well. While the managements of educational institutions significantly rely on the use of this communication platform, satisfying the information needs of the student population has to be based on an analysis of user habits and behaviour according to the routes supplied by the links located on the respective web pages. The collection of user accessibility patterns in such divided information-based environment is called accessibility route pattern mining. The respective examination process has to cope with certain difficulties including the excessive amount of data found on the web, the complexity of the given web page, the dynamic changes of the respective web pages, and the anomalies concerning the identification of the users.

The research effort utilises web log mining as one of the branches of web mining. The web mining is a quantitative methodology of examination based on primary data.

Web mining is one of the sub fields of data mining, the selection, retrieval or extraction of knowledge from large amounts of information held in data bases, data storage facilities, or other information repositories (Han-Kamber, 2004). Thus through web mining we can obtain new knowledge inaccessible by the use of existing data base technology devices. Data mining can be used in all areas in which a large amount of data is produced including business, public administration, health care, education, etc. While data mining is based on statistics, its use of advance mathematics and informatics far exceeds the limits of the statistical approach (Bodon, 2006).

Web mining examinations and analyses focus on web accessibility patterns, web structures, and the regularity and dynamics of web contents. Web mining efforts entail three main areas: web content mining, web structure mining, and web usage mining. However, since web structure is part of web content, we can restrict the original three examination areas to two: namely web content mining and web usage mining (Han-Kamber, 2004).

Our present analysis primarily focuses on web usage mining based on web log mining. During the examination of the log entries we attempt to disclose and identify the accessibility patterns of web pages in order to improve the quality of the respective web-based services or that of the capacity of the web servers. A given web server prepares a record concerning the accessed web page, and the collection of the respective accessibility data is called web log, which is stored in a record file, better known as log file. While the content of the given web log entry might vary, each web server retains the following data:

- The IP address of the source of the given data request,
- The requested URL address,
- and some additional data.

Web log data, however, cannot be processed immediately as preliminary processing facilitated

by the Web mining CAT module is required for assuring the validity and reliability of the respective information. The preliminary processing effort including purification, condensation, and transformation is followed by analysis revealing the browsing habits of the users.

Preparatory steps of the examination process

The research effort performed at the Human Computer Interaction Research Group (HCI) of the Ergonomics and Psychology Department of the Institute of Applied Pedagogy and Psychology (IAPP) at the Budapest University of Technology (BUT) was based on the use of the data mining software product of the SPSS Clementine along with the pertaining Web mining CAT module.

The HCI research group is allowed to use the costly, but indispensable web mining software free of charge by the help of the SPSS Hungary until the end of 2007 exclusively for non-profit purposes. In addition to the data mining Clementine Software the Web Mining CAT module facilitating the practical aspects of web mining supported the research process as well.

Starting in September 2006, being a part of the Professor Lajos Izsó-led HCI Research Group of the Ergonomics and Psychology Department of the IAPP at BUT I was given an opportunity to become familiar with the theoretical and practical foundations of web mining technologies. Consequently, the basis for the research efforts described in this presentation was provided by my participation in the research group, and the implementation process was supported by the monthly-held team meetings.

For the purpose of the relevant examinations, the system manager of the web server of the Eszterházy Károly College provided the access log files needed at various times of the research effort. As a starting point of the research effort assisted by the CAT module of the Web mining program we surveyed and assessed the number of the most popular events.

Event Category	Event Name	Event Definition	Event Attributes
Homepage	Homepage	/	
Homepage	Kezdooldal	/index.php	
Keresesek	Keresesek	/kereso.php/*	
Telefonkonyv	Telefonkonyv	/telefonkonyv/*	
Tanulmányi_ugy	Tanulmányi_ugy	/tik/*	
Sajatsite	Sajatsite	/~bota1/*	
Mas_kepzes	Mas_kepzes	/oktatas/*	
EKF_egysegek	EKF_egysegek	/szervezet/*	

Figure 3. Event file prepared for the research effort.

Events can be defined as the beginning of downloading by clicking (sending or registering) onto a link, menu, picture, or button presented on the browsing surface. The number of events is significant for establishing the frequency of web page visits and in order to limit data processing time events forming the focus of the research effort should be prioritised. The respective event file is basically a text-based document presented in Figure 3.

While an event file can contain many events, due to temporal and spatial restrictions we will focus only on a small part of the events. Although we have to fill in the first two columns, this cannot be done without a thorough knowledge of the examined web pages, and in the third column the exact access route of the references stored in the log file has to be entered. The Instruction menu point was renamed Other Type of Training (MAS kepzes), while web pages prepared by instructors were given the Own Page (Sajatsite) event name. Figure 4 uses the former name.

The results of the first examination

In comparison to the approximately 200,000 total hits registered on the EKC web page, the Instruction menu point is quite frequently chosen as the menu and the respective components recorded more than 11,000 hits in January and more than 8,000 hits in February.

Although in January more than 500 times education-related web pages were hit before reaching the Instruction menu, this does not necessarily prove that the given menu point is misleading. However, web pages chosen after hitting "Instruction" are more illustrative of the problem as more than 650 hits out of 5000 events, or choosing the given web page, indicated that the visitor requested more Instruction-related information. Furthermore, the Organisational Units (Szervezeti egységek) menu point displaying available training programs is frequently visited by the users of the College web page as well.

The selected two-part or dual routes better illustrate the problem. 4.21 % of the visitors decided to return to the opening page (step back), 1.99% hit the Organisational Units menu point, and 0.63 % opted for the search function of the web page. Although these numbers appear negligible compared to the total two-part routes, this actually means 5343 clicks. Furthermore, 3.50% of the total number of three-part route selections indicated that following two hits users returned to the opening page (step back), 1.36 % selected the Organisational Units menu, which means 3353 clicks. In case of 1.20% of all four-part hits during which the users consequently hit on the Instruction and Organisational Units menu points 1086 clicks were registered. Consequently, having examined the number of identical series of clicks, it can be concluded that the number of clicks between the Instruction, Organisational Units, and the home page is rather significant.

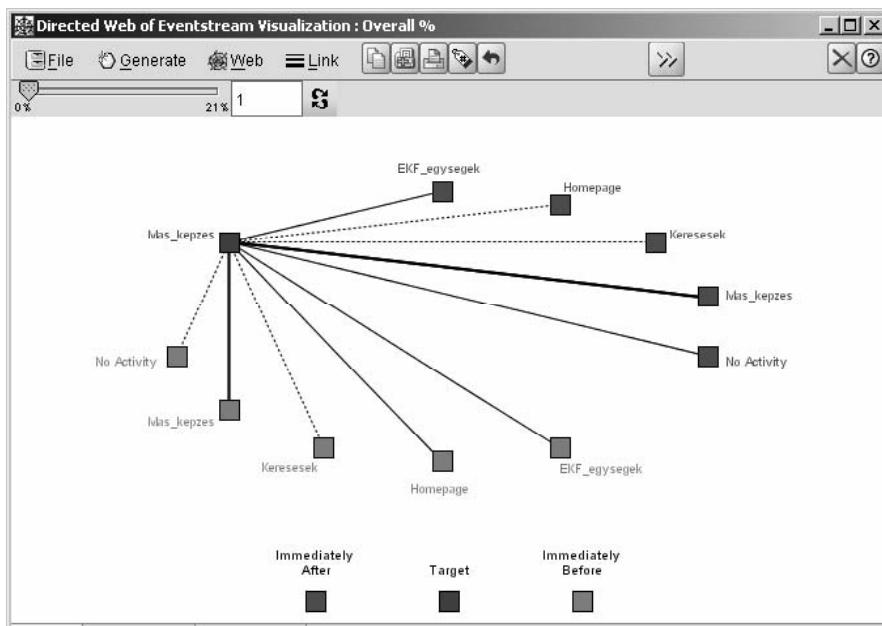


Figure 4. Web pages (events) reached before and after hitting the Instruction (Mas_kepzes) menu.

Having examined events hit after the Instruction menu point we can conclude that several users get lost within the system as one of the largest amounts of clicks refers to Organisational Units (EKC units) indicated by a relatively bold printed, not yet broken line. While the users expect access to such aspects and features of the College as departments, study units, and web pages dealing with academic matters, the menu points do not help them in realising this goal.

The web page has a few dead ends as well as the disappointed user returns to a certain web page (in case of the present examination it is the home page) with a few clicks and terminates the visit. Whereas the home page provides a continuous display of the menu items, 135 return clicks were registered from the Instruction menu point indicating the termination of the respective visits. Although this fact would not prove the misleading potential of the Instruction menu point, along with the abovementioned developments it can be proven that the naming of the respective menu point is deceptive for many users and results in misdirected search efforts.

I have informed the developers of the home page of this quality shortcoming and they will

take my research into consideration during the elaboration of the new web site. Nevertheless, the respective examinations and analyses should be continued while using the data concerning the old and new web pages.

The results of the second examination

Whereas navigation between the different sites is mostly supported by local references, users tend to avoid structural links and prefer search engines except when in need of in-depth processing of information located on the given web-page. Management however insists on the use of menu points as the web pages cannot be contextualised without structural links (Kiss, 2006).

The design principles of ergonomic experts take the working memory capacity of the respective visitors into consideration (Shneiderman, 1987, 1992). According to Shneiderman and other researchers (1989: Nilesen, Molich; 1988: Roe; 1987: Gardiner, Christie) information combinations whose size exceeds the recommended 7 ± 2 cognitive scheme can lead to time-consuming, mentally strenuous processing and user disorientation in addition to raising the potential of diminishing user satisfaction.



Figure 5. Menu points accessed after the selection of a sub menu.

The new webpage has 23 menu points located on the left panel in addition to 7 menu points located in the upper section complemented by additional menu points in English and Hungarian along with the menu points of the home page of the city of Eger. The above lay-out also includes the *Back to Main Page* and the *Back* menu points (Figure 5). The menus are divided into subsections and this somewhat reduces the orientation capability. At any rate the number of the menu points is significant. Our examination reveals that a certain group of menu points have hardly been selected. While the violation of the previously discussed guideline and the resulting disorientation or false information providing capability might provide an explanation, researchers must take into consideration the data reflecting a limited, that is a only a few weeks, time frame. Consequently, the use of certain menus are connected with a different period in the given year thus in order to substantiate our hypothesis we need to review a greater time frame of the respective weblog file in a repeated examination.

The misleading appearance of the Instruction menu revealed during the first inquiry has been eliminated. The surveying of consecutive series of clicks showed that visitors did not follow the consecutive and unnecessary navigation routes connecting the various menu points. As the inquiry revealed the most frequently visited menu points focus on a variety of topics. The *Webcamera* menu point received more than 40,000 clicks, the College's on-line registration and recording system, the

Neptun menu was hit more than 4,500 times, the menu point prepared for *Prospective Students* was visited 3,800 times, the submenu of the *Registrar and Academic Information System* was clicked around 3,000 times, the search feature of the home page was clicked 2,900 times, the *Phonebook* menu point received approximately 2,500 hits and the *Library* menu point was clicked more than 1,500 times. Our examination revealed that the menu point prepared to inform prospective students (*Prospective Students*) despite the high number of clicks was placed rather low in the list ranking 17th in the left side menu. Moreover, while the *Library* menu point posted a high visitor frequency, it was not included even in the second level submenus.

In case of the top section, after clicking either on the *Museums* menu point or hitting its sub-points new information can be accessed, while clicking on the *Services* menu does not provide any information as new knowledge can only be accessed from the submenus. Visitors tend to be interested mostly in the *Museums* main menu point providing contact information concerning one of the most frequently visited museums in the country. Furthermore, the submenu points primarily provide historical descriptions. Nevertheless, the examination reinforced that this menu is the least deceiving as with the exception of the *Services* menu point the other menus have no subcomponents.



Figure 6. The lay-out of the *Museums* and *Services* menu points after being selected by users.

Among the submenus of the *Introduction* menu point located in the left panel of the new home page (Figure 6) the *Faculty Structure* submenu point has not been selected at all. This can be due to the easier route provided by the graphic icons of the 4 organisational units facilitating a separate availability route to the four faculties (Humanities, Social Science and Economics, Natural Sciences, Teacher Training and Knowledge Technology). Visitors prefer such means of availability which provide structurally easier access with the least possible mental strain (Figure 7).



Figure 7. The lay-out of the *Introduction* menu point after being selected by users.

However, the *Faculty Structure* submenu point presents the same information as the *Training Structure* menu point. It is reasonable to conclude that one link was not connected appropriately and the reference was not corrected during the copying of the source code of the home page. The misinformation potential is increased as information concerning the faculties and the respective training schemes is available at two other locations in the two submenus of the *Organisational Units* menu point which does not have any submenus. It can be argued that the increased availability of the same information despite the partially differing content is not necessary. The examinations focusing on longer intervals might reveal the misinformation potential of menu clusters.

Summary

The present analysis has proven that even a most carefully designed home page can mislead the respective visitor. Naturally a large section of misinformation-related errors is not due to careless design, but it reflects unique, unidentified on-line behaviour and habits of users. We believe that these examinations should continue and after the identification of ergonomic shortcomings the inquiries should focus on user personalisation and eventually home page development.

The use of web mining software provided useful tools for leaders of educational institutions to deal with these problems. While the above examinations supplying accurate data concerning every visitor assist managers, the leaders of educational institutions have to draw the respective conclusions and identify the respective enhancement trends with the development team of the given home page.

This option can provide reliable and correct examination data in order to support management decisions. The greatest obstacle in its wide-spread application is the required financial and intellectual commitment. It is expected that the price of the program will slowly decrease and a free access home page using Web 2.0 technology can facilitate broader use for management purposes.

References

- Bodon, F. (2006). *Adatbányászati algoritmusok. Data mining Algorithms. Electronic document.* URL: <http://www.cs.bme.hu/~bodon/magyar/adatbanyaszat/tanulmany/index.html>
- Han, J., & Kamber, M. (2004). *Adatbányászat. Koncepciók és technikák. Data mining: Concepts and Technologies.* Budapest: Panem.
- Izsó, L. (2006). *A webbányászat alapjai. The Foundations of Webmining. Electronic document.* Budapest. URL: <http://www.erg.bme.hu/webmining/>
- Kiss, O. E. (2006). Információ-keresés a hipertext alapú környezetek szerkezetének és a felajánlott navigációs eszközöknek a függvényében. Information Search in the Context of the Structures of Hypertext-based Environments and of the Available Navigational Devices (Doctoral dissertation, University of Eötvös Loránd). Budapest.
- Shneiderman, B. (1987, 1992). *Designing the User Interface.* Reading, MA. Addison-Wesley.
- SPSS Inc. (2005). *WebMining for Clementine 1.5 Application template.* Chicago, USA.
- SPSS Inc. (2005). *WebMining for Clementine 1.5 User's Guide.* SPSS Inc. Chicago, USA.

*Advised by Janis Gedrovics,
Riga Teacher Training and Educational Management Academy, Latvia*

László Bóta

PhD. student at Eszterházy Károly College, Eger, Hungary.
Eszterházy tér 1. H-3300 Eger, Hungary.
Phone: 36/36-520-400.
E-mail: botal@ektf.hu
Website: <http://www.ektf.hu/english/index.html>