PROBLEMS
OF MANAGEMENT
IN THE 21st CENTURY
Volume 1, 2011
26

SUPPORTING E-MARKETING DECISION MAKING BY THE MANAGEMENT OF THE ESZTERHÁZY KÁROLY COLLEGE VIA BEHAVIOUR-BASED SEGMENTATION OF THE VISITORS OF THE INSTITUTIONAL WEB-PAGE

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

Abstract

The increasingly competitive higher educational environment compels the management of universities and colleges to assign high priority to an overall maximisation of client services. Consequently, while academic leaders must become familiar with the aspects of on-line communication much favoured by today's younger generation, the intensification and improvement of the quality of available on-line services cannot be imagined without reliable information on the Internet use habits and behaviour of clients.

The managers and administrators of Hungarian college and university websites are mostly unfamiliar with the web-related conduct or habits of their customers as in case of long-running web-pages based on an unchanging structure only basic visitor statistics are available at best. Yet marketing communication decisions should be based on information reflecting real website-consumer traits acquired via a more professional analysis. Data mining is one such decision-making support mechanism.

Data mining models are capable of revealing and predicting information hidden beneath the respective critical mass. Therefore inspired by the methodology of marketing science this type of research concentrates on the segmentation of on-line consumers via the elaboration of visitor clusters.

The present article provides a scientific overview and analysis of the main difficulties related to cluster construction, especially the development of the relevant algorithmic forms. The successful application of the model provides much-needed reliable and vital support to the institutional decision making process. Thus pertinent data yielded by cluster research can facilitate more effective on-line service customized to the needs of the users.

Key words: *clustering model, data mining, marketing communication, on-line conduct, web-ergonomics*.

Introduction

As a result of the potential elimination and integration of universities and colleges along with reductions in the governmentally supported student population the already intensive competition in the Hungarian higher education sphere is expected to intensify in the near future. Consequently, successful and efficient strategic marketing activities are vital for the the long term survival of colleges and universities (Töröcsik & Kuráth, 2010).

Marketing communication and especially Public Relations activities are integral components of any strategic marketing activity (Kotler & Fox, 1995, p.356). The results of the present research effort whose primary objective is the improvement of the image and the respective

PROBLEMS
OF MANAGEMENT
IN THE 21st CENTURY
Volume 1, 2011

attitudes towards the Eszterházy Károly College via the examination of its main PR device, (the institutional home page) are expected to support marketing communication decisions made by institutional management. The forthcoming inquiry utilizing data mining cluster models focuses on the on-line conduct of the targeted public and aims to contribute to a more personalized customer service and information provision process as well. Moreover, the examination based upon a variety of web-ergonomic considerations will include recommendations for the improvement of the respective electronic surfaces too.

While data mining methods have been deployed with large, profit-oriented business firms for many years, such approach has been adopted by the higher education sphere only since 2007. Whereas most inquiries focus on the on-line conduct and habits of students utilising e-learning surfaces, only a few researchers recommended the use of data mining models for the facilitation of institutional managerial decision making (Balogh & Horváth, 2010). Consequently, based upon the data gained from the on-line administration surface attempts have been made to segment the student population according to strategic marketing aspects. To the researcher's best knowledge, however, no data mining-based analysis of web-pages maintained by a higher education institution have been performed until now, thus the research to be described below can be considered an unprecedented and pioneering endeavour in Hungary.

Data Mining as a Methodology and Research Tool

Data mining has enjoyed an increasing significance as a research method since the 1990s. This approach entails an iterative process during which intelligent manoeuvres or operational sequences are performed in order to identify data patterns. Intelligent operations imply various statistics-based analytical techniques and methods including neural networks, factor analysis, and cluster analysis (Bodon, 2010).

The present paper utilizing a variety of professional terminology examines human web use from various vantage points. Consequently, the web user is considered a consumer from a marketing stand point, while according to web-ergonomical or web-mining considerations (s)he is categorised as the visitor or user. Such concepts are deployed in an overlapping manner as the respective terms are considered synonyms of each other.

The Introduction of Data Mining and the Description of the Applied Program

In addition to the fields of telecommunication and medical sciences data mining efforts can help in the realization of such business-related goals as the assessment of potential credit risks, the analysis of credit applications, marketing oriented classification and clustering of consumers, investigation of financial crimes, the examination of the efficiency of advertising campaigns, and the retention of consumers (Han & Kamber, 2004, p.447).

Data mining efforts can be grouped into two categories. Descriptive data mining reveals the general features of data while the predictive version anticipates, makes inferences, or prognoses from the available data. Web mining is one of the subfields of the descriptive data mining category. Web mining focuses on web accessibility patterns and web structures in addition to the regularity and dynamics of web content. Moreover, since web structures are part of web content web mining also examines web content and web use mining (Han & Kamber, 2004, p.433). Web content mining plays a significant role in marketing research, helps the mapping and exploring of markets, facilitates the development of pricing policies, and contributes to the selection of distribution channels, along with the elaboration of an appropriate communication strategy via the analysis of the on-line appearance of competitors. Our research, however, focuses on web use mining exploring the habits and conduct patterns of consumers thereby fulfilling crucial market development purposes. Consequently, web use mining can be also be

considered web log mining as it is based on the examination of web log entries compiled by web servers.

Web log mining efforts use a wide variety of key terms. An event refers to an user's specific request for downloading a webpage, document, or image carried out during the given on-line visit. Visit refers to a limited series of requests originating from an user. A visit is considered completed if the initial request is not followed by another within 30 minutes. If the logfile includes another request after 30 minutes, that is considered a new visit related to the same user.

In the context of the present research the user is unidentified as his or her activity can only be traced according to the IP address (Figure 3: Host name) as (s)he did not have to register with an user name or pass word and no cookie was assigned to his or her computer either.

Presently there are two leading data mining program packages available worldwide: IBM's SPSS modeler program and the Enterprise Miner program of SAS. Both packages are complemented by the webmining apparatus as well. These products, however, due to their prohibitive cost are not available to higher education institutions. The data mining software used for the present research effort: the IBM SPSS Modeler 14.1 and the earlier version of Web Mining for Clementine 1.5 Application Template (CAT) (SPSS, 2005a), was provided cost-free by SPSS Hungary for the non-profit research efforts of the Budapest University of Technology and Economics.

The Introduction and Description of the Clustering Process

Clustering has been the most often applied aspect of data mining. While clustering can entail a wide variety of areas including the grouping of web-pages, genes, diseases, and clients, personalized service by categorizing and differentially treating the resulting groups of clients and consumers has witnessed the most dynamic development. The main reason for clustering is the costs associated with the manual categorization of a large number of clients. From both marketing and research aspects the primary focus in not on the categorization effort or the allocation of the respective persons into given categories themselves, but on the shared characteristics of the specific groups (Bodon, 2010, p.147).

Clustering, unlike most typical categorization efforts refers to grouping or segmentation without pre-determined criteria. The main aim of clustering is to separate similar and different components into varying groups. One of the chief difficulties associated with the production of an algorhythm facilitating appropriate or unequivocal group formation is determining what can be considered the main feature of the given category as even college students can be grouped according to various criteria. This problem, however, as demonstrated by its application in marketing research can be eliminated by automatization (Bodon, 2010, p.147).

The IBM SPSS Modeler software offers three cluster facilitating algorhythms: Kohonen, K-Means, and the TwoStep (SPSS, 2009b). While the overall research effort requires the identification of the most appropriate one, the primary focus of the present essay is on the algorhythm that had proven to be most effective.

The Applied Data Mining Methodology

The present research was based on the CRISP-DM (CRoss Industry Standard Process for Data Mining) approach whose flow chart is described by Figure 1. While this method was elaborated by SPSS and other leading representatives of the industry in 1996, SAS has also developed its own methodology known by the acronym of SEMMA (Sample, Explore, Modify, Model, Access). This latter one, however, tends to emphasize the technological elements associated with data mining. The 6 steps of the CRISP-DM embody the life cycle of a data mining

PROBLEMS
OF MANAGEMENT
IN THE 21st CENTURY
Volume 1, 2011

project (Abonyi, 2006, p.19). Furthermore, most data mining tasks are of the iterative nature requiring the multiple performance of the various inquiry components accompanied with the respective modifications.

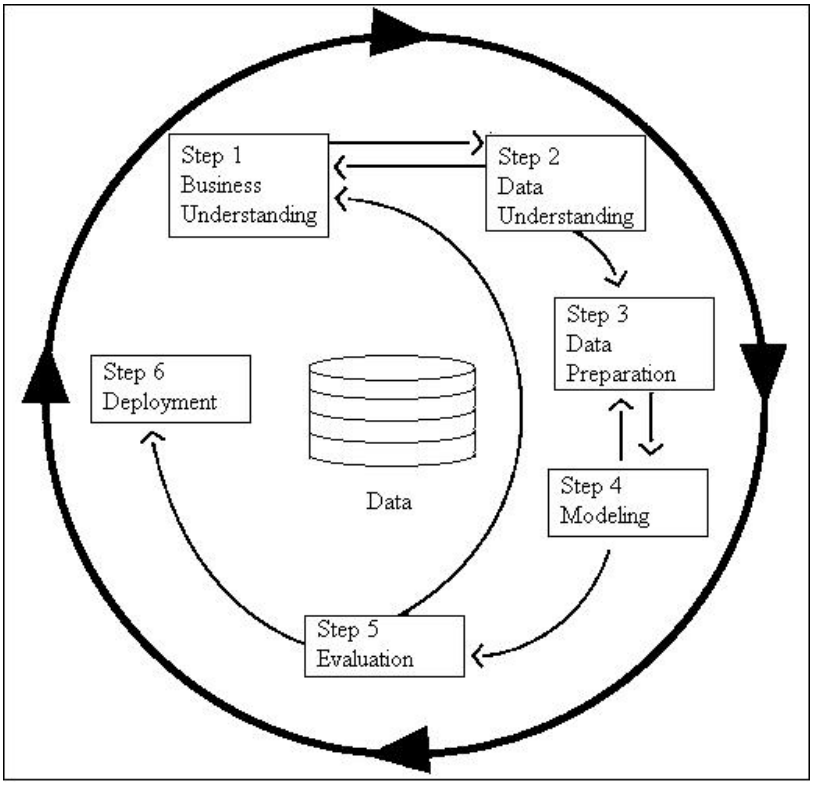

Figure 1: Phases of the CRISP-DM Process Model (The CRISP-DM consortium, 2000).

The Main Objective and Phases of the Research Effort According to the CRISP-DM Approach

Hypothesis

The primary purpose of the research coincides with a business-oriented goal as well. Namely, having examined the selection patterns of the main menus of the institutional home page groups of users with identical choices should be compiled. The establishment of the clusters and the subsequent web-ergonomic evaluation can determine whether the positioning of the respective menu points promotes or hinders the members of the particular group in navigating on the home page. In the present context global navigation implies the set of virtually all menu points accessible at the given home page. In order to carry out the aforementioned task we resort to cluster-facilitating algorthythms.

Consequently, the following hypothesis is put forth: in case of home pages consisting of static web pages the application of the results gained from data mining clustering models can lead to the improvement of the efficiency of the services provided for unidentified on-line visitors.

Step 1: The definition and explanation of the business-related objective

The resulting segmentation should facilitate the improvement of marketing communication while providing services meeting personal needs. Consequently, navigation on the home page should be improved to an extent perceivable by the members of the established strategically important segments. Since a large segment of the student population of the College is below 30 and extremely proficient in the use of on-line surfaces the resulting service improvement can go a long way in retaining them as clients or consumers.

The professional or research objective is the rephrasing of the business application goal namely the segmentation of clients according to the respective visits and global navigation activities. Futhermore, based upon the segments incorporating the exploration of particular behaviour profiles the operation of the menu structures should become more effective from a web-ergonomical point of view as well.

Step 2: The examination of the available data

The previous home page of the Eszterházy Károly College (http://www.ektf.hu) operated until 2007 and it was replaced by a new one as of October 9 of the same year at the same URL address. While the original version is not available anymore, the data examined in this essay were preserved in the respective weblog files. Moreover, although the data were registered in the weblog file from January 7, 2007 until the date of the home page conversion, not all pertaining information could be used for the purpose of the present research.

The effectiveness of the research was somewhat limited by the fact that the users were not identified. The home page of the College primarily fulfils an information provision function concerning the availability of instructors, entrance requirements, and course descriptions etc, thus in most cases the users or consumers do not register. While the on-line visits of unindentified consumers can only be made relevant if the user started his visit from the same IP address, most Hungarian service providers issue dynamic IP labels resulting in the use of a new IP address after repeated signing on or after a certain duration of time (one week) expires. In our case this is less problematic since we are only interested in the clustering of visitors according to particular features of conduct.

PROBLEMS
OF MANAGEMENT
IN THE 21st CENTURY
Volume 1, 2011

Figure 2: The global navigation of the examined web page. The left side has 14 menu points, the right side has 21 menu points and 5 further auxiliary menu points are found at the upper section in addition to the respective search fields.

The data mining software makes the presentation of statistic data possible as well. The software was able to process data obtained in the period between Jauary 7, 2007 and March 18, 2007. Accordingly the home page (Figure 2) was visited by 67,837 users (out of which only 88 originated from the premises of the College) realizing 183,283 visits. The most frequently used menu point was the *Information for Future Students* with 919,744 hits, second place was taken by *Organizational Units* with 103,882 hits, while the *Instruction* menu located on the left side of the webpage came in third with 28,854 hits.

Step 3: The preparation of the data

This phase of the data mining effort is called data purification and data transformation. Data in original form are not suitable for the carrying out of the examination and the preparation is a multi-step process which helps the model to produce the clusters. The starting data include the set of information recorded in the weblogfile. In order to perform this operation we use the User Mode Determination Stream of the Web Mining CAT then we adapt the elements of the stream to the given task. The stream is the execution program of the given inquiry whose steps, also called nodes, are designated by symbols.

The first step of data purification is the selection of relevant records by the Web Mining node (Figure 3). The input of the node is the logfile, the output is the fields described below:

1. Event ID, 2. Event Category, 3. Event Name, 4. Resource (URL of event), 5. Event

Timestamp, 6. Visit ID, 7. Visit Start Timestamp 8. User ID, 9. User Type, 10. Authorized User Name, 11. User Cookie, 12. Hostname, 13. Attribute ID, 14. Attribute Name, 15. Attribute Value (SPSS, 2005b, p.8).

Event ID	Event Category	Event Name	Resource	Event Timestamp	Visit ID	Visit Start Timestamp	User ID	Us...	Au...	Us...	Hostname	Attribute ID	Attr...	Attr...
1	Homepage	Homepage	/	20070204 06:36:09	14	20070204 06:36:09	14	3			:80.99.18.138	0		
2	Szervezet	Foisk_Karok	/szervezet/karok.html	20070204 06:36:15	14	20070204 06:36:09	14	3			:80.99.18.138	0		
3	Webcam	Webcam	/kamera/indexcam.php	20070204 06:38:54	16	20070204 06:38:54	16	3			:81.182.124.5	0		
4	Webcam	Webcam	/kamera/indexcam.php	20070204 06:40:08	16	20070204 06:38:54	16	3			:81.182.124.5	0		
5	EKF_egysegek	EKF_egysegek	/szervezet/intezetek/gazdasagtud.htm	20070204 06:40:41	18	20070204 06:40:41	18	3			:72.30.252.173	0		
6	Homepage	Homepage	/	20070204 06:42:14	22	20070204 06:42:14	22	3			:80.98.48.21	0		
7	Szervezet	TIK	/szervezet/tik	20070204 06:42:39	22	20070204 06:42:14	22	3			:80.98.48.21	0		
8	Tanulmanyi_ugy	TIK	/tik	20070204 06:42:56	23	20070204 06:42:56	23	3			:74.6.86.82	0		

Figure 3: The output fields of the Web Mining node.

The second, more complex step of data purification is the aggregation of data leading to a five part record structure including such components as visit identifier, visitor identifier, event identifier, event name, and the number of hits. If an user chooses more than 4 menu points, that information is not included in the research effort as the multiple choice points more to the uncertainty and misorientation of the user, than to a conscious selection. (Figure 4)

Figure 4: Nodes represent the data preparation process in the web mining software stream.

Eventually a crossreference chart is prepared recording the connection of the user and the visit with the chosen menu point that is the event (Figure 5).

	User ID	Visit ID	Event Name_Admin	Event Name_Alapitvany	Event Name_Allasborze	Event Name_Alt_info	Event Name_BaratiKor	Event Name_Bels
11260	14	11176	F	F	F	F	F	F
11261	14	14	F	F	F	F	F	F
11262	13	36131	F	F	F	F	F	F
11263	13	25767	F	F	F	F	F	F
11264	13	25315	F	F	F	F	F	F
11265	11	20973	F	F	F	F	F	F
11266	11	20443	F	F	F	F	F	F
11267	11	16898	F	F	F	F	F	F

Figure 5: The frequency of events during the respective user visits (Event Name_xxx) The meaning of the respective field contents: T=true (visit took place), F=false (no visit took place).

Step 4: The construction of the model

The production of the model begins with the Auto Cluster mode facilitating the comparison of three clustering algorythms. While Figure 6 and especially the Silhouette column suggests that the Kohonen model would be the most suitable for this purpose, 44 clusters appear to be too much for the grouping of visitors.

Since the home page consists of 41 menus the construction and application of the TwoStep model containing fewer clusters can provide an adequate research methodology. Consequently, the 6 cluster TwoStep model reflecting virtually the same capabilities as that of the Kohonen model was chosen.

PROBLEMS
OF MANAGEMENT
IN THE 21st CENTURY
Volume 1, 2011

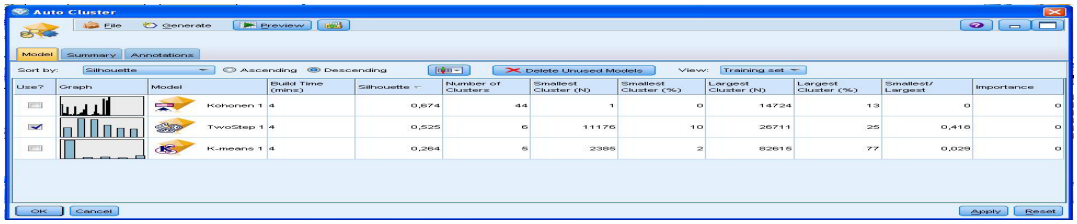

Figure 6: The comparison of the features of the three clustering models.

The TwoStep node uses a two-step clustering method. The first step makes a single pass through the data to compress the raw input data into a manageable set of clusters. The second step uses a hierarchical clustering method to progressively merge the subclusters into larger and larger clusters. TwoStep has the advantage of automatically estimating the optimal number of clusters for the training data. It can handle mixed field types and large datasets efficiently (SPSS, 2009a, p.387).

After the learning or experimental phase the applied version of the program presents the usable construct immediately. Figure 7 contains data characteristic of the cluster compiled according to this model. The quality of the model is registered in the lower regions of the Good domain, an acceptable value in itself.

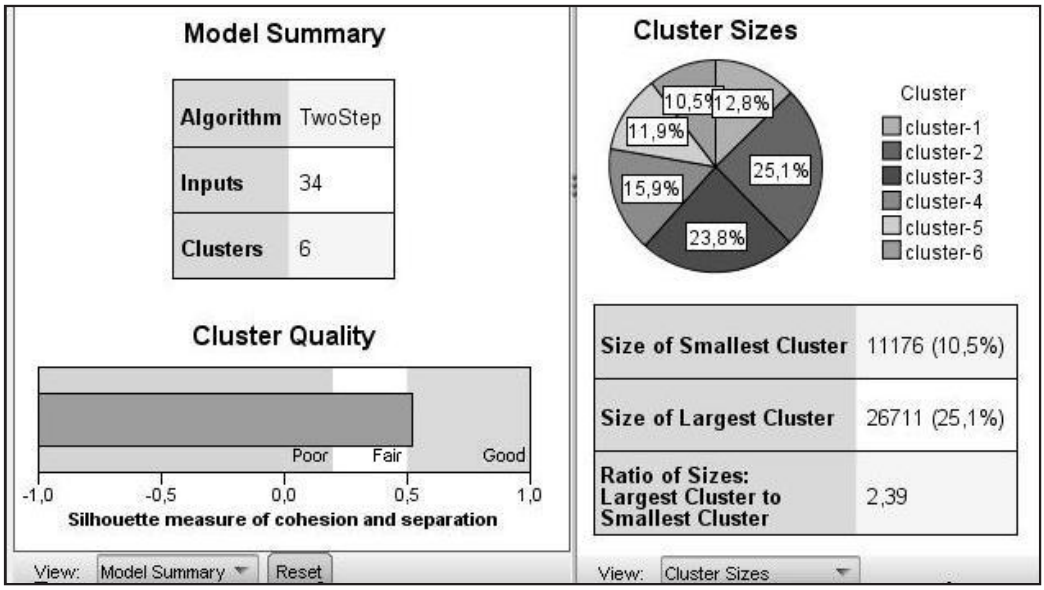

Figure 7: The comparison of the clusters produced by the TwoStep Cluster.

Having correlated the meanings with the clusters the user segments reflecting navigation activities are produced:

- cluster 1 (12.8%): visitors preferring the web camera option
- cluster 2 (25.1%): visitors searching for general information
- cluster 3 (23.8%): visitors interested in the structure of the College
- cluster 4 (15.9%): visitors searching for information concerning registration or general administration-related information from the Academic and Student Information Centre
- cluster 5 (11.9%): NEPTUN (online academic grade registration system)- users

- cluster 6 (10.5%): Future students of the College interested in entrance examination-related information

Results of Research

Figure 8 reveals the detailed analysis of the correlation of the menu system with the respective segments according to web-ergonomical considerations. Visitors favouring the web camera (cluster-1, 12.8%) can access the service by clicking on the picture located below the menu on the left. While the number of downloads is significant, its PR potential is considered average as its location in the lower region of the left side menu is merely satisfactory.

Visitors in search of general information (cluster-2, 25.1%) are grouped into a set reflecting a wide variety of visits not belonging to any other sets.

Visitors interested in the structure of the College (cluster-3, 23.8%) select the *Faculties, Organisational Units* or *Dormitories* menu points. The menu displaying information on the Faculties is located on the top of the right side column of menu groups. Placing this information on the left is not justified either by its content or the hit number as the most crucial menu points should be located in that section. The *Dormitories* menu point is located at the bottom of the left side menu groups. However the *Faculties, Organisational Units* or *Dormitories* menu points (Figure 1) appear to violate the disjunctivity principle of menu design as the same units are accessible via several menu points. Moreover, the *Organisation Units* menu point appears not to meet the principle of totality either as it does not contain information on all units. Furthermore, the placement of menu points in separate menu groups should be reconsidered as such arrangement might pose additional difficulty for the users.

While the segment analysis on the *Academic and Student Information Centre* (cluster-4, 15.9%) reveals a considerable hit ratio this menupoint is not integrated either in the left or upper menu sets preferred by users as it is part of the right side menu group. Nevertheless, the significance and importance of this organisational unit calls for easier accessibility.

Where as the Neptun system (cluster-5, 11.9%) is accessible via the eigth point of the menu group on the left, there is another identical menu point in the upper section as well. The placement of the two menupoints in such distance is unnecessary and redundant.

The cluster group of entrance applicants (cluster-6, 10.5%) tend to select the *Entrance Requirements, Academic and Student Information Centre* and *Instruction* menus available at the 9th place on the right side, on the 3d place on the right side, and as the 5th menu point on the left respectively. Consequently, the connection of the entrance requirements with the information on the instruction activity at the College can mean additional mental burden for those interested in such information as the *Instruction* menu point contains data not fully relevant to its name as it primarily focuses on complementary training programs. Moreover, these menu points relevant to future students are located at a considerable distance from each other and one is named in a rather misleading way.

PROBLEMS
OF MANAGEMENT
IN THE 21st CENTURY
Volume 1, 2011

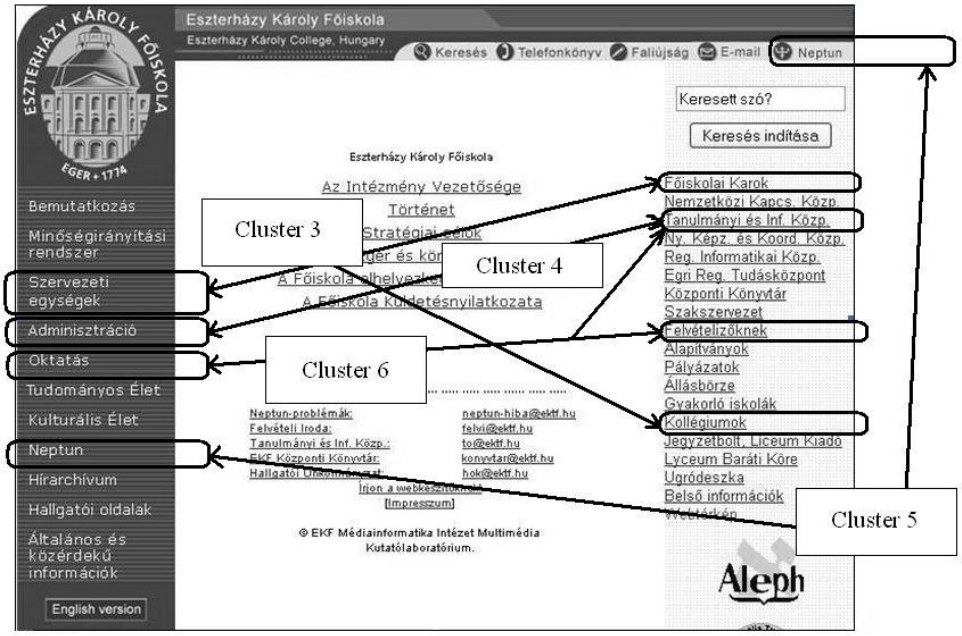

Figure 8: Menu points of the College homepage related to clusters 3, 4, 5, and 6.

In light of the above we can conclude that the hypothesis of the research is substantiated as web-mining efforts based upon cluster models can help in the improvement of the efficiency of the on-line services of the College home page.

Step 5: The business-related evaluation of the results

The evaluation of the quality of user support provided by the home page is an additional goal of the research program. Step 5 sums up the main results of the business-oriented quality assessment of support services available to the respective segments. Web-ergonomic considerations suggest that users tend to avoid difficult to navigate web sites and in case of additional problems they discontinue the visit (Krug, 2008, p.21). The present research revealed numerous web-ergonomic problems relevant to specified segments which would have been impossible to do without the formation of groups. Consequently, the disclosure of numerous obstacles associated with the navigation of the respective web-page warrant the re-consideration of the e-marketing related decision making of institutional management.

Discussion

The conduct of user groups is only partially supported by the home page in question. Specific menupoints used by respective segments are located far from each other, thus the modification or the reconsideration of the menu structure is recommended. Consequently, the management of the institution has the following choices: accept the recommendations and initiate the change of the menu structure of the web page, expand the inquiry onto the deeper level of the menu structure including the exploration of 2d and 3d level menus, perform a more detailed analysis of the second cluster, or identify the frequently and consecutively used menu points. While, the web page of the College has been fully restructured via the compilation of

varying segments relevant to the new surface, global navigation can still be examined. The other direction of the examination focuses on the additional electronic surfaces used by the College facilitating the promotion of targeted and specific marketing efforts.

Steps 6: The business-oriented application of said results

It has been proven that it is possible and worthwhile to improve navigation features as the users can be retained on the long run this way. A consumer representing the Net generation will be more loyal to the institution if (s)he can find the required information without any difficulty.

In 2007 management further motivated by the identification of additional problems and short-comings decided to authorize the elaboration of a fully different home page. The College's home page with a renewed look and structure is still in use today.

Conclusions

While not in possession of the respective research results, the management of the Eszterházy Károly College has made the right decision concerning the total restructuring of the institutional web-page. Although familiarity with the specified data could have lead to the required changes earlier, quick and scientifically sound decision-making along with satisfying consumer demands via e-marketing methods has helped to maintain a competitive edge. The research effort revealed the advantages obtained via the clustering of the visitors of electronic surfaces based upon the recording of user conduct information.

While clustering data related to unidentified users can lead to the improvement of the respective services, the clustering of electronic surfaces in case of identified users can be recommended as well. Regarding e-learning programs web mining can provide information concerning students' learning habits and the respective methodological background, while the exploration of the student registration system can help the elaboration of an institutional marketing strategy. All in all, the analysis of the respective segments has exposed correlations indispensable for long term business-related decision making.

References

Abonyi, J. (Ed.) (2006). *Adatbányászat a hatékonyság eszköze. Gyakorlati útmutató kezdőknek és haladóknak.* Budapest (Hungary), Computerbooks.

Balogh, I., Horváth, Á. (Eds.) (2010). *Felhasználói viselkedés elemzése webes környezetekben. Webanalitikai módszerek alkalmazása viselkedés-elemzésre.* Budapest (Hungary), DSGI Mérnöki Iroda.

Bodon, F. (2010). *Adatbányászati algoritmusok.* Budapest (Hungary). [online technical book] Retrieved January 20, 2011, from Budapest Univercity of Technology and Economics website: http://www.cs.bme.hu/~bodon/magyar/adatbanyaszat/tanulmany/adatbanyaszat.pdf

The CRISP-DM consortium (2000). *Cross Industry Standard Process for Data Mining.* CRISP-DM Consortium, [online document]. Retrieved April 19, 2011, from http://www.crisp-dm.org

Han, J., Kamber, M. (2004). *Adatbányászat. Koncepciók és technikák.* Budapest (Hungary), Panem.

Kotler, P., Fox, K. F. A. (1995). *Strategic Marketing for Educational Institutions.* Englewood Cliffs, New Jersey (USA), Prentice-Hall Inc.

Krug, S. (2008). *Ne törd a fejem! Felhasználóbarát webdizájn.* Budapest (Hungary), HVG. [The Hungarian translation based on: Steve Krug (2006). Don't Make Me Think! A Common Sense Approach to Web Usability. Second Edition. California (USA), New Riders Publishing Berkeley.]

SPSS (2005a). *Web Mining for Clementine 1.5 Application Template.* Chicago (USA), SPSS Inc. [electronic document].

SPSS (2005b). *Web Mining for Clementine 1.5 User's Guide.* Chicago (USA), SPSS Inc. [electronic document].

SPSS (2009a). *PASW Modeler 13 Modeling Nodes.* Chicago (USA), SPSS Inc. [electronic document].

SPSS (2009b). *PASW Modeler 13 User's Guide.* Chicago (USA), SPSS Inc. [electronic document].

Töröcsik, M., Kuráth, G. (Eds.) (2010). *Egyetemi marketing. Marketing a felsőoktatásban. [University marketing. Marketing in higher education.]* Pécsi Tudományegyetem, Pécs, Hungary.

Advised by Vincentas Lamanauskas, University of Siauliai, Lithuania

Received: *May 15, 2011* Accepted: *May 28, 2011*

László Bóta Ph. D. Student, Assistant Professor, Eszterházy Károly Főiskola (Eszterházy Károly College), Institute of Mediainformatics, Department of Informatics, H-3300 Eger, Eszterházy tér 1, Hungary.
E-mail: botal@ektf.hu
Website: http://www.ektf.hu/~botal