Personalized Web Advertising Method

Przemyslaw Kazienko[1], Michal Adamski[1,2]

[1] Wroclaw University of Technology, Department of Information Systems,
Wybrzeze S. Wyspianskiego 27, 50-370 Wroclaw, Poland
kazienko@pwr.wroc.pl, http://www.pwr.wroc.pl/~kazienko
[2] Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia,
Monte de Caparica, 2829-516 Caparica, Portugal
michal.adamski@iol.pt

Abstract. Personalization of online advertising is a great challenge while the market is moving and adapting to the realities of the Internet. Many existing approaches to advertisement recommendation are based on demographic targeting or on information gained directly from the user. In this paper we introduce the AD ROSA system for automatic web banner personalization, which integrates web usage and content mining techniques to reduce user input and to respect the user's privacy. Furthermore, the advertising campaign policy, an important factor for both the publisher and advertiser, is taken into consideration. To enable online personalized advertising the integration of all relevant information is performed in one vector space.

1 Introduction

In the age of disappearing borders and mixed societies the current demographic targeting of freely available Web content seems to be insufficient. The market consists of human beings, not demographics, so web personalization should depend on an individual's behavior rather than on stereotypes created according to his or her geographical location or other demographic features (e.g. gender, age). Traditional advertising serving the same offers for everyone does not meet the current requirements of businesses. To increase the effectiveness, the right person should receive the right message at the right time and in the right context [1].

Web advertising is mainly done with banners – graphical elements on a web page, or with their 'mutations' – displayed in a new layer or new window of the browser. There are many other forms of online advertisements like sponsored links or articles, or mail-outs, but in this article we will only concentrate on banners and similar forms.

Users, showered with hundreds of advertisements, often pay less attention to banners appearing on a web page as bitmap images or animations, and this seems to be the main problem of web advertising. The solution is to increase the correspondence between user interests and the subject of the displayed advertisement [4].

Two significant research domains may be distinguished within Internet advertising: scheduling and personalization. The main goal of the former is to maximize the total click-through-rate for all advertisements by appropriately managing of exposition time and advertising space on the web page [3, 14].

The latter seems to be an important and difficult challenge for current advertisers. It aims to assign a suitable advertisement to the user, so it is necessary to have some information about the user. Many web portals create user profiles using the information gained during the registration process or ask the user to answer some questions about their preferences. However, this requires a lot of time and effort, and that can discourage many users. Besides, users tend to give incorrect data when being concerned for their privacy [13]. Even reliable data becomes out-of-date with the evolution of the online customer's interests. An alternative solution is to exploit information stored in the web server logs. This method is safe in regard to privacy fears and may also be useful for news portals or web sites where users do not need to log in to use the service [19]. Another approach to advertisement personalization is presented in [11]. Short-term and long-term interests of the user were identified. Short-term interests are derived from the keywords submitted by the user in searching services. However, such keywords may often have nothing in common with the user's regular preferences. Long-term interests are proposed to be taken from user profiles, which are completed by users and stored in the database of the system. However, advertising personalization was performed using only short-term information.

A system based on web usage mining and clustering of navigation paths to create usage patterns was presented in [19]. Pages from both the publisher's web site and the advertisements' target sites are manually classified into thematic categories by experts. The assignment of appropriate advertisements to each active user is accomplished according to pages (categories) visited by the given user during the current session. This matching is based on fuzzy rules stored in the system. The fuzzy approach was also used in target advertising based on user profiles [21].

Three main advertising models can be distinguished [5]: broker, portal and advertiser models. In the *broker model* there exists an advertising broker that connects publishers (web sites in which advertisements would be displayed) and advertisers (companies providing banners to be emitted). The broker often provides some targeting options. This model is applied i.e. by DoubleClick [15]. The *portal model* (used by large web portals) is the special case of the broker model: the publisher owns the advertisement management software and cooperates with many advertisers. The *advertiser model*, in which the advertiser manages the advertisements, allows big online store to display its banners on pages of particular portals.

2 Advertisement Features

Nowadays most online advertising systems use the principle of *the customer-based targeting*. Each user is identified and classified according to his or her geographical location (IP address) and browser settings sent with the HTTP request, navigation habits and user profiles (preferences) completed by the user during the registration process. This data is used to personalize the displayed banner advertisement [2, 11].

Analyzing advertising offers of the greatest Polish portals (*www.wp.pl*, *www.onet.pl*), we observed many target criteria available for advertisers. Apart from the demographic data of a user (age, gender, location, etc.), advertisements can be targeted towards the user's education, profession or interests. Furthermore, the publisher can choose the time of day of the emission, particular parts of the web page, and limit the number of emissions for a single user. An advertiser is usually charged on the basis of cost per month per one thousand emissions of advertisement (CPM). Another approach is the usage of click-though-rate (CTR) - the ratio of the number of clicks to the emission number [2]. It should be mentioned that the average CTR is currently decreasing as a consequence of the increasing number of total advertisements displayed [17].

The method presented below takes into consideration most of the contemporary applied aspects of advertising campaigns with respect to users' privacy rights.

3 AD ROSA System - Method Overview

The advertising method proposed in this paper solves the problem of automatic personalization of web banner advertisements with respect to user privacy (none of the user's personal details are stored in a database) and recent advertising campaign policy. It is based on knowledge extraction from the web pages' content and historical user sessions as well as the current behavior of the online user, using data mining techniques. The implementation of data mining to web content and web usage is usually called web content and web usage mining, respectively [12, 20]. There are also some integration methods of both these approaches [8, 9, 12].

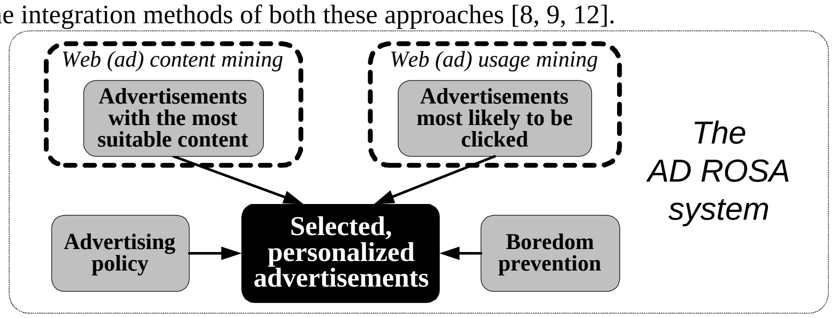

Fig. 1. Factors of advertisement selection in the AD ROSA system

The proposed method uses both web mining techniques and combines in one personalized framework several useful factors of advertising: the most suitable content (the content of the advertiser's web site), click probability, advertising policy i.e. arising from contracts and boredom prevention mechanisms (Fig. 1). The latest, are responsible for periodical rotation – the scheduling of advertisements for the user. The AD ROSA system is part of the ROSA project [6, 7, 8, 10].

Historical user sessions are stored in the database and clustered to obtain typical, aggregated user sessions (Fig. 2). The cluster's centroid corresponds to one *usage pattern* of the publisher's web site. Each user session is linked up to the set of advertisements visited (clicked) by the user during this session (*visited ad vector*). Having a cluster of sessions, the AD ROSA system can also extract information about related, visited advertisements, by counting the mean vector (*ad visiting pattern vector*) from all visited vectors related to sessions from the cluster. Thus, one web usage pattern (centroid) corresponds to exactly one ad visiting pattern (centroid).

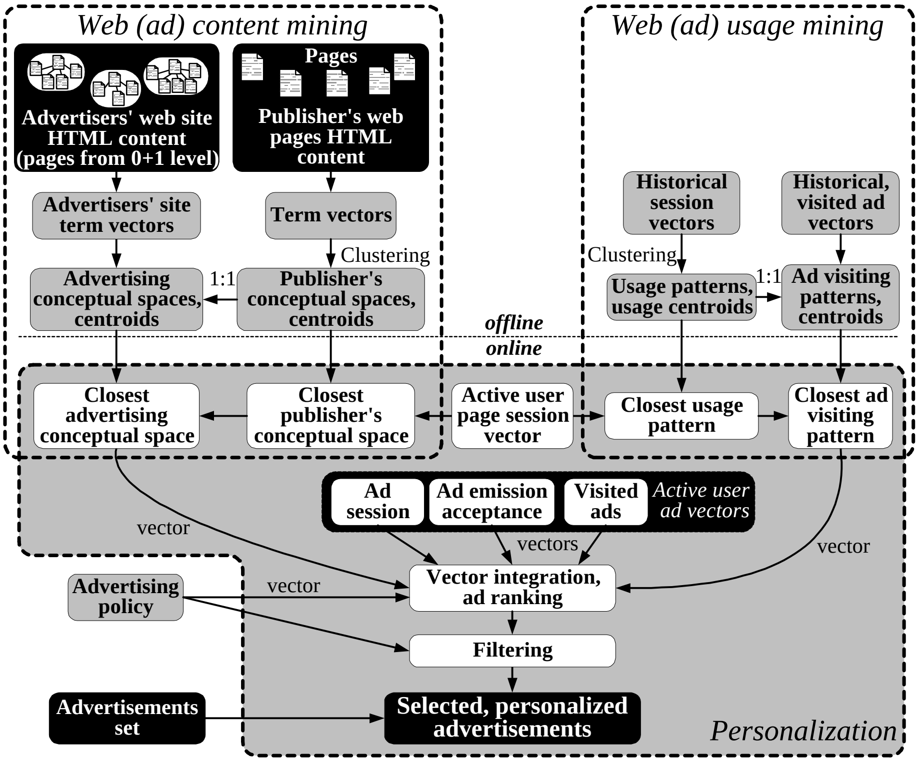

Fig. 2. The overview of the personalized advertising method in the AD ROSA system

The site content of the publisher's web pages is automatically processed in the similar way. Content thematic groups - *conceptual spaces* - are received using the clustering of term vectors extracted from the HTML content of web pages [8, 9].

In order to recommend a suitable advertisement for the user we have to know its general subject matter. This is achieved by text (HTML) content analysis of the advertisement target web site. The AD ROSA system automatically downloads advertiser's web pages and processes only the terms, which occur in the publisher's web pages. As a result we obtain *advertising conceptual spaces* corresponding to the appropriate *publisher's conceptual spaces*.

A user requesting a web page is assigned to both the closest usage pattern and the closest conceptual space, based on the user's previous behavior during the session - the *active user page session vector*. Assuming that he or she will behave like others (usage pattern) and is interested in web pages similar to those recently visited (conceptual space), the AD ROSA system can recommend to the user the most suitable advertisements. Advertisements related to the closest conceptual space link to web sites with content appropriate for the current user, so they should be displayed to the user. The closest usage pattern and in consequence the closest ad visiting pattern enables the selection of advertisements that are most likely to be clicked by the user.

User behavior (*active user session vector*) as well as information about already displayed or visited banners (*active user ad session* and *active user ad visited vector*) are stored by the system separately for each user. They all prevent the too frequent emission of the same advertisement for one user and provide control over the number of emissions contracted in the advertising campaign. A new, personalized advertisement ranking is determined for the user after each user's request. This process exploits information mentioned above and the targeting parameters established by the advertiser (*advertising policy*), like limited emissions per user during a single session (*active user ad emission acceptance vector*). As a further advantage for the publisher, some additional priority features of each advertisement can be set up manually (*advertising policy*). This provides an opportunity to increase the ranking's value for more profitable advertisements.

Finally, the personalized ranking list is filtered using additional advertising policy features: limitation to certain web browsers, time of day of the emission, etc. In the end the AD ROSA system returns to the web server the list of *n* top ranked, filtered advertisements that are dynamically incorporated into the returned web page content.

4 Content Processing - Web Content Mining

The publisher's web content is processed using *crawler* - an agent that downloads and indexes the content of all pages from the web site [10]. Terms obtained from HTML content are filtered using several statistical features to extract best descriptors [9]. For each selected term t_i an M-dimensional *term page vector* $\boldsymbol{tp_j} = < w_{j1}^{tp}, w_{j2}^{tp}, ..., w_{jM}^{tp} >$ is created. The coordinate w_{ji}^{tp} denotes the weight of the term t_j in the document (page) d_i according to Information Retrieval theory [18]:

$$w_{ji}^{tp} = tf_{ji} * \log\left(\frac{M}{n^{t_j}}\right) \tag{1}$$

where: M – the number of pages in the whole publisher's web site, tf_{ji} - term frequency (the number of occurrences) of the term t_j in the page d_i, and n^{t_j} – the number of pages in which the term t_j occurs.

The set of $\boldsymbol{tp_j}$ vectors is clustered, using the group average link – a hierarchical agglomerative clustering method (HACM) - to discover groups of terms that are close to each other [16]. Applying this method to a selected web site with ca. 3100 pages and 500 filtered descriptors, 41 clusters were obtained [9].

Terms from one cluster describe the *publisher's conceptual spaces* (thematic groups) existing within the publisher's web site. Once we have clusters we can calculate essences of *conceptual spaces* – centroid (mean) vectors - as follows:

$$\boldsymbol{ctp_k} = \frac{1}{max_k} \sum_{l=1}^{n_k} \boldsymbol{tp_{lk}} \tag{2}$$

where $\boldsymbol{ctp_k}$– the centroid of the k-th cluster; $\boldsymbol{tp_{lk}}$ – the l-th term vector belonging to the k-th cluster; n_k – the number of terms in the k-th cluster; max_k – the maximum value of the sum of component coordinates in the k-th cluster, used for normalization [8].

The content of the target web site of an advertisement is processed similarly. A typical banner links to the main page of the target service (level 0), which often includes just a menu or is the redirection page. For that reason the AD ROSA system also analyzes all pages from the next level (level 1) – pages from the same domain linked to the level 0 page. All pages from levels 0 and 1 are concatenated and treated by the system as the advertiser's content.

For each term extracted from target pages, which simultaneously exist in the set of term page vectors, *the advertiser term vector* $\boldsymbol{ta_j} = < w^{ta}_{j1}, w^{ta}_{j2}, ..., w^{ta}_{jN} >$ is created; where N – the number of advertisements (web sites). The coordinate w^{ta}_{ji} denotes the weight of the term t_j in the advertiser's web site (a_i) and is calculated using (1). Please note that one advertiser term vector $\boldsymbol{ta_j}$ corresponds to exactly one publisher's term page vector $\boldsymbol{tp_j}$. For that reason terms from publisher's web pages that do not occur in any advertiser's web site have the vector $\boldsymbol{ta_j}$ with all coordinates set to zero. This ensures a uniform term domain for both the publisher's and the advertiser's content.

Advertiser term vectors $\boldsymbol{ta_j}$ are not clustered because the equivalent publisher's term page vectors have already been clustered. Since one vector $\boldsymbol{ta_j}$ corresponds to one vector $\boldsymbol{tp_j}$, one publisher's conceptual space is equivalent to one *advertising conceptual space*. Only one mean vector - centroid $\boldsymbol{cta_k}$ - for each k-th *advertising conceptual space* is calculated.

5 Session and Clicked Advertisement Processing - Usage Mining

The first step of usage mining is the acquisition of HTTP requests and the extraction of sessions. A user session is a series of pages requested by the user during one visit to the publisher's web site. Since web server logs do not provide any easy methods of grouping these requests into sessions, each request coming to the web server should be captured and assigned to a particular session using a unique identifier passed to a client's browser [8]. Each j-th user session stored by the system is represented by the M-dimensional *session vector* $\boldsymbol{s_j} = < w^{s}_{j1}, w^{s}_{j2}, ..., w^{s}_{jM} >$; where $w^{s}_{ji} \in \{0,1\}$ denotes whether the i-th page was visited (1) or not (0) during the j-th session.

Historical session vectors $\boldsymbol{s_j}$ are clustered into K' separated usage clusters in the same way as term page vectors $\boldsymbol{tp_j}$. The centroid $\boldsymbol{cs_k}$ of such a cluster (*usage pattern*) describes one typical user's behavior - the navigation path throughout the web site. For an example web site with over 7700 sessions (35000 requests) 19 clusters were created [9]. Please note that coordinates of the centroid $\boldsymbol{cs_k}$ belong to the range [0,1].

Data about visited (clicked) advertisements during the j-th user session is stored in *the visited ad vectors* $\boldsymbol{v_j} = < w^{v}_{j1}, w^{v}_{j2}, ..., w^{v}_{jN} >$; w^{v}_{j1} is the number of clicks of the i-th advertisement during the j-th session. For each user session $\boldsymbol{s}$, there exists exactly one corresponding *visited ad vector* $\boldsymbol{v}$. Thus, having the k'-th session cluster $\boldsymbol{cs_k}$ we also obtain the appropriate cluster of *visited ad vectors* without clustering procedure - similarly to *the publisher's* and *the advertising conceptual spaces*. For each k'-th cluster the centroid - *visiting pattern* $\boldsymbol{cv_{k'}} = < w^{cv}_{k'1}, w^{cv}_{k'2}, ..., w^{cv}_{k'N} >$, $w^{cv}_{k'i} \in [0,1]$, is found:

$$\boldsymbol{cv_{k'}} = \frac{1}{max_{k'}} \sum_{l=1}^{n_{k'}} \boldsymbol{v_{lk'}}, \; max_{k'} > 0, \tag{3}$$

where $n_{k'}$ – the number of vectors in the k'-th cluster, $max_{k'}$ – the maximum, aggregated value of visits of a single advertisements in the k'-th cluster: $max_{k'} = \max_{i=1,2,...N}\left(\sum_{l=1}^{n_{k'}} w_{li}^{v}\right)$. The assumption $max_{k'}$ >0 means there must be at least one visit to any advertisement in the cluster. Otherwise all coordinates of $\boldsymbol{cv_{k'}}$ are set to 0. Note that for the most often visited advertisement in the cluster $w_{k'i}^{cv}=1$.

6 Active User Monitoring

The behavior of each active user visiting the publisher's web site is monitored until the end of the user's session. The AD ROSA system keeps the information about documents visited by all active users. For the j-th active user the *page session vector* $\boldsymbol{ps_j} = < w_{j1}^{ps}, w_{j2}^{ps}, ..., w_{jM}^{ps} >$ is maintained; where $w_{ji}^{ps} \in [0,1]$ denotes the importance (timeliness) of the i-th page for j-th active user:

$$w_{ji}^{ps} \begin{cases} (?)^{n_{ji}^{ps}}, & \text{when document } d_i \text{ was visited during the } j\text{-th active session} \\ 0, & \text{when document } d_i \text{ was not visited during the } j\text{-th active session} \end{cases} \tag{4}$$

where: ? — the constant parameter for the interval [0,1], determined experimentally, in the implementation ?=0,95 was assumed; n_{ji}^{ps} — the consecutive index of the document d_i in the j-th active session in reverse order. For the just viewed document $n_{ji}^{ps}=0$ ($w_{ji}^{ps}=1$), for the previous document $n_{ji}^{ps}=1$ ($w_{ji}^{ps}=?\le 1$), etc. If the document was visited more than once, the lowest value is assumed to n_{ji}^{ps} [9].

The active user ad session vector $\boldsymbol{as_j} = < w_{j1}^{as}, w_{j2}^{as}, ..., w_{jN}^{as} >$ plays a similar role to the page session vector in relation to displayed advertisements. It prevents advertisements from being displayed too often and enables their periodical rotation. The coordinate $w_{ji}^{as} \in [0,1]$ denotes when the i-th advertisement was displayed to the j-th current user. Values in the vector are always updated after advertisements have been assigned to the user and displayed on the web page. The w_{ji}^{as} value is set to 1 for the just emitted i-th advertisement after the j-th user's request. At the same time all other w_{ji}^{as} values are decreased using factor $a \in [0,1]$, as follows:

$$w_{ji}^{as} = a * w_{ji}^{as} \,. \tag{5}$$

It was assumed in the implementation that a=0.8. *The active user ad session vector* $\boldsymbol{as_j}$ with value zero at all positions is created with the first request from the active user and is removed after the user's session has finished.

Information about the number of emissions of every advertisement is stored in *the ad emission vector* $\boldsymbol{e} = < w_{j1}^{e}, w_{j2}^{e}, ..., w_{jN}^{e} >$; where value of w_{ji}^{e} is the number of emissions of the i-th advertisement for the j-th active user. Information kept in *the ad* *emission vector* is necessary in order not to display one advertisement too many times to one user, and is useful in controlling advertising policy.

7 Advertising Policy

A lot of publishers allow limiting emission of one advertisement to a user during a single user's session. The number of permitted emissions of the *i*-th advertisement is denoted by the coordinate w_i^{epu} of *the emission per user vector* ***epu***$=< w_1^{epu}, w_2^{epu}, ..., w_N^{epu} >$. The information about the acceptance of the emission for the *j*-th active user the *i*-th advertisement is stored in *the active user ad emission acceptance vector* $\boldsymbol{uea_j}=< w_{j1}^{uea}, w_{j2}^{uea}, ..., w_{jN}^{uea} >$. The values of its coordinates depend on the general limit of emissions (***epu***) and the current number of emissions of the *i*-th advertisement to the *j*-th active user (***e***), as follows:

$$w_{ji}^{uea} = \begin{cases} 1, & \text{if } w_i^{epu} - w_{ji}^{e} > 0 \quad \text{or} \quad \text{"emission is unlimited"} \\ 0, & \text{otherwise} \end{cases} . \tag{6}$$

As mentioned above, the publisher has the ability to increase the importance of each advertisement. The adequate, manually set priorities are stored in *the ad priority vector* ***p***$=< w_1^p, w_2^p, ..., w_N^p >$, where $w_i^p \in [0,1]$.

8 Vector Integration, Personalized Ranking, Filtering

At each HTTP request from the *j*-th user the AD ROSA system again assigns to this user, described by $\boldsymbol{ps_j}$ (see section 6), the closest *publisher's conceptual space* ($\boldsymbol{ctp_k}$) and the closest *usage pattern* ($\boldsymbol{cs_{k'}}$), searching for centroids with the minimum value of $cos(\boldsymbol{ps_j},\boldsymbol{ctp_k})$ and $cos(\boldsymbol{ps_j},\boldsymbol{cs_{k'}})$, respectively. Each *publisher's conceptual space* $\boldsymbol{ctp_k}$ corresponds to one *advertising conceptual space* $\boldsymbol{cta_k}$ and each *usage pattern* $\boldsymbol{cs_{k'}}$ is related to one *visiting pattern* $\boldsymbol{cv_{k'}}$. In consequence, we obtain $\boldsymbol{cta_k}$ and $\boldsymbol{cv_{k'}}$ suitable for the current behavior ($\boldsymbol{ps_j}$) of the *j*-th active user.

Having obtained all the above-mentioned vectors, the personalized advertisement ranking is created for each user: the list of the most appropriate advertisements is obtained by sorting coordinates of the *rank vector* - $\boldsymbol{rank_j}$. This vector integrates all the *N*-dimensional vectors engaged into the personalization process:

$$\boldsymbol{rank_j} = (1-\boldsymbol{v_j}) \otimes (1-\boldsymbol{as_j}) \otimes \boldsymbol{uea_j} \otimes \boldsymbol{p} \otimes (\boldsymbol{cta_k}+\boldsymbol{cv_{k'}}+b), \tag{7}$$

Operator ⊗, used for two vectors, denotes the multiplication of individual coordinates of these vectors: the *i*-th coordinate of the first vector is multiplied by the *i*-th coordinate of the second vector, $i=1,2,\ldots,N$. This produces the third vector with the same dimension.

The closest *advertising conceptual space* and *visited pattern* may have all coordinates equal to 0, when similar users have not visited any advertisements and the terms from the closest *publisher's conceptual space* do not occur in any advertiser's web sites. For that reason, the constant $ß$=0,5 was introduced. It enables new advertisements to be recommended, even through they could have null values in $\boldsymbol{cta_k}$ and $\boldsymbol{cv_{k'}}$.

Ranking vector includes all information useful for recommendation. Owing to 1–$\boldsymbol{v_j}$ banners clicked by the current user are omitted, while 1–$\boldsymbol{as_j}$ prevents individual advertisements from being exposed too often for one user. $\boldsymbol{uea_j}$ is responsible for monitoring whether the limit of advertisements per one user has been reached and $\boldsymbol{p}$ respects manually specified priorities. $\boldsymbol{cv_{k'}}$ is used in order to encourage the display of advertisements that have been clicked by users who visited similar web pages as the current user. Similarly, the use of $\boldsymbol{cta_k}$ promotes the display of advertisements linking to web sites that contain similar words to the pages previously visited by the current user.

Next, the ordered list of advertisements is filtered using additional advertising policy features stored in the database. In this way the requirements of certain web browsers or the time of day of the emission can be fulfilled. All advertisements are also filtered according to their shape, strictly determined by the page layout. As a result, the AD ROSA system delivers personalized, periodically changed advertisements meeting various advertising policy features.

9 Conclusions and Future Work

The method of advertising personalization presented in this paper integrates information coming from different sources: web usage mining, web content mining, advertising policy and boredom prevention. The large number of considered factors means that the same user on the same page may each time be recommended different advertisements. All processes in the method (Fig. 2) are performed automatically by the system, which decreases management costs. The idea of personalization based on "user-friendly" data acquisition (without the user's effort) makes the AD ROSA system applicable in almost any open-access, anonymous web portals and can widen demographic personalization systems used in many web sites. The integration of the AD ROSA system with the ROSA core systems, which recommends hyperlinks, results in the complex personalization system satisfying both users and advertisers.

Future work will concentrate on the optimization of online processes and the development of an advertisement scheduling system, which is an important issue when dealing with many advertisers. In e-commerce, the method can be extended to include purchases history and product ratings gathered by the system.

References

1. Adams R.: Intelligent advertising. AI & Society, Vol.18, No.1 (2004) 68 - 81.
2. Aggarwal C.C., Wolf J.L., Yu P.S.: A Framework for the Optimizing of WWW Advertising. Trends in Distributed Systems for Electronic Commerce, International IFIP/GI Working Conference, TREC'98, Hamburg, Germany, LNCS 1402, Springer Verlag (1998) 1-10.
3. Amiri A., Menon S.: Efficient Scheduling of Internet Banner Advertisements. ACM Transactions on Internet Technology, Vol. 3, No. 4 (2003) 334–346.
4. Baudisch P., Leopold D.: User-configurable advertising profiles applied to Web page banners. Proc. of the First Berlin Economics Workshop, Berlin, Germany (1997)

http://patrickbaudisch.com/publications/1997-Baudisch-Berlin-UserConfigurableAdvertisingProfiles.pdf.
5. Bilhev G., Marston D.: Personalised advertising – exploiting the distributed user profile. BT Technology Journal, Vol. 21, Issue 1 (2003) 84-90.
6. Kazienko P.: Multi-agent Web Recommendation Method Based on Indirect Association Rules. KES'2004, 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems, Wellington, New Zealand, LNAI, Springer Verlag (2004).
7. Kazienko P., Kiewra M.: Integration of Relational Databases and Web Site Content for Product and Page Recommendation. 8th International Database Engineering & Applications Symposium. IDEAS '04, Coimbra, Portugal, IEEE Computer Society (2004).
8. Kazienko P., Kiewra M.: Link Recommendation Method Based on Web Content and Usage Mining. New Trends in Intelligent Information Processing and Web Mining Proc. of the International IIS: IIPWM´03 Conference, Advances in Soft Computing, Springer Verlag (2003) 529-534.
9. Kazienko P., Kiewra M.: Personalized Recommendation of Web Pages. Chapter 10 in: Nguyen T. (ed.) Intelligent Technologies for Inconsistent Knowledge Processing. Advanced Knowledge International, Adelaide, South Australia (2004) pp. 163-183.
10. Kazienko P., Kiewra M.: ROSA - Multi-agent System for Web Services Personalization. AWIC'03. First Atlantic Web Intelligence Conference Proceedings, LNAI 2663, Springer Verlag (2003) 297-306.
11. Langheinrich M., Nakamura A., Abe N., Kamba T., Koseki Y.: Unintrusive Customization Techniques for Web Advertising. Computer Networks, Vol. 31 (11-16) (1999) 1259-1272.
12. Mobasher B., Dai H., Luo T., Sun Y., Zhu J.: Integrating Web Usage and Content Mining for More Effective Personalization. EC-Web 2000, LNCS 1875, Springer (2000) 156-176.
13. Montaner M., López B., de la Rosa J. L.: A Taxonomy of Recommender Agents on the Internet, Artificial Intelligence Review, Kluwer Academic Pub. Vol. 19 (4) (2003) 285-330.
14. Nakamura A.: Improvements in Practical Aspects of Optimally Scheduling Web Advertising. Proceedings of the 11th Int. WWW Conference, WWW2002, ACM Press, (2002) 536-541 http://www2002.org/CDROM/refereed/295/.
15. Online Advertising. DoubleClick Inc. (2004) http://www.doubleclick.com/us/products/online_advertising/
16. Rasmussen E.: Clustering algorithms. In: Frakes W., Baeza-Yates R. (eds): Information retrieval: data, structures & algorithms. Chap. 16, Englewood Cliffs, NJ, Prentice Hall (1992) 419-442.
17. Rodgers Z.: Volume Up, Click-Throughs Down in Q4 '03 Serving Report. Jupitermedia Corporation, February 5 (2004) http://www.clickz.com/stats/markets/advertising/article.php/3309271.
18. Salton G.: Automatic Text Processing. The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley, Reading, MA (1989).
19. Sung Min Bae, Sang Chan Park, Sung Ho Ha: Fuzzy Web Ad Selector Based on Web Usage Mining. IEEE Intelligent Systems, Vol. 18, No. 6, November/December (2003) 62-69.
20. Yao Y.Y., Hamilton H.J., Wang X.: PagePrompter: An Intelligent Agent for Web Navigation Created Using Data Mining Techniques. RSCTC 2002, LNCS 2475 Springer Verlag (2002) 506-513.
21. Yager R.R.: Targeted E-commerce Marketing Using Fuzzy Intelligent Agents. IEEE Intelligent Systems, Vol. 15. No. 6, November/December (2000) 42-45.