Beyond user clicks: an algorithm and an agent-based architecture to discover user behavior

E. Menasalvas (1†), S. Millán (2), M. Pérez (1), E. Hochsztain (3*), V. Robles (1), O.Marbán (1), J. Peña (1), A. Tasistro (3)

(1) Facultad de Informática UPM. Madrid Spain (2) Universidad del Valle. Cali. Colombia.
(3) Facultad de Ingeniería - Universidad ORT Uruguay

Abstract. One of the gaps that has to be filled for the fully deployment of web activities is a closer human relationship with the user. This lack can only be dealt with understanding user online actions, which will make proactive and autonomous web behavior possible. In order to find out what navigators behavior looks like it is crucial to understand the semantics underlying their clicks. Being aware of current navigation patterns is not always enough. Knowledge about user objectives and session goals can be discovered from the information collected and tracked by web clicks. Nevertheless, information on these clicks has to be enriched with semantics if user behavior understanding is the ultimate goal. This paper presents an approach that provides an estimation of the end result of a user navigation in terms of site goals achievements. Once clickstream information has been enriched we propose to apply a method based on discriminant analysis in order to obtain two different results: (i) the relevant factors that contribute most to the success of a session and (ii) a statistical classification method to estimate the result of an ongoing session. In order to carry out this proposal, we also present the design of an agent-based architecture in which the role of each agent is deeply analyzed.

1. Introduction and related works

The Web is not only a technology to connect computers; it is also a new way of communication, cheaper and with greater location independency than the traditional one. This explains the amazing number of organizations that during the last decade have started their traditional activities in the Internet, designing and implementing web sites to interact with their customers.

It is possible to say that the web is becoming one of the main communication channels for any kind of transaction, being commerce obviously one of its main uses. However, web-based activities loose direct contact with clients and, therefore, fail to achieve many of the features that enable small businesses to develop a warm human relationship with customers. This must not be understood as if those features cannot

[+] Research is partially supported by Universidad Politécnica de Madrid (project Web-RT)
[*] Research has been partially supported by Programa de Desarrollo Tecnológico (Uruguay).

be translated onto the web. In this sense, it is important to say that many enterprises have been very worried about getting hold of the identity of the navigator. In the traditional way of making business, a good CRM means providing a user with the right product at the right moment. What is important is not the identity of the user but his or her likes and dislikes, his or her preferences, the way he or she behaves. Thus, trying to have a good e-CRM, the behavior of the user has to be analyzed and for doing so the only available data are the clickstream. In this sense, most of enterprises are investing great amounts of money establishing mechanisms to discover internet's user behavior.

Based on data mining techniques many approaches [9,16,21,23,25,29,31,34,35,38] have been proposed for tracking and analyzing clickstream data in order to obtain most frequent paths. Most of them calculate user profiles taking this information into account. Nonetheless, just knowing most frequent patterns is not enough; it is of vital importance to combine this information with company goals with the intention of making this process more profitable for web site sponsors.

Consequently, the challenge is to get, analyze and understand the behavior of every user that connects to the web site in order to improve information on the webhouse, make better decisions and take action. Based on annotations and ontologies, several methods and approaches have been proposed in order to take these site semantics into account [13, 26, 5, 17, 36, 19, 10, 33, 6]. A framework for web personalization that integrates domain ontologies and usage patterns is presented in [13]. Semantic information could be obtained from textual content included in the site or using conceptual hierarchies based on services [7]. An approach that takes into account associated information goals inferred from particular patterns and from information scent associated to linked pages is proposed in [11]. In [33] a site's pages are classified according to their function: action and target pages. On the other hand, in [26] user actions are represented based on an ontology's taxonomy. URL's are mapped to applications events depending on what they represent (actions or content).

In [20] we propose an algorithm that takes into account both the information of the server logs and the business goals improving traditional web analysis. In this algorithm, however, the value of the links is statically assigned. Although this approach is useful for finding the value of the path the user has taken, this value depends on the value of the links. In this paper, we propose a method based on discriminant analysis that makes it possible to establish factors that are relevant for determining the success or failure of a session with respect to business goals. Then based on these factors and applying a statistical classification method we provide the site administrator with a predictive model to estimate the result of an on-going session.

Considering that we can have dynamic pages, the algorithm does not take the page or link the user has visited as parameters. It takes instead certain concepts (e.g., factors, actions) extracted from them that are interesting for the site.

The key point in these approaches is how company goals are translated into parameters and values to be automatically managed during web server performance. These values are required to be able to predict the most valuable (in terms of the site goal) ongoing sessions. The different paths themselves are not the focus – what is

important is to discover, given a visited sequence of pages, the final result of the session and we propose to estimate this result based not on the pages the user has visited but on the semantics underlying these pages.
The point here to be emphasized is that the proposed approach is very effective as we are dealing with a predictive algorithm whose model though it has been calculated off-line by analyzing past sessions, needs to be online to be applied. The predictive model not only estimates the result of the session but also identifies the relevant factors that make a session successful. With these values, we provide the site administrator with enhanced information to find out which action should be performed next in order to provide the best service to navigators and to be more competitive. This information can be applied as an important parameter for web server load balancing, priority scheduling or even web server acceptance. Many other performance and tuning parameters can be affected by this goal-oriented analysis.

In spite of the huge volume of data stored in the web, the relationship between user navigation data and site effectiveness, in terms of site goals when trying to design "good pages" from the site users' point of view, is still difficult to understand. Several approaches, models and measures [8, 34, 3, 15, 22] have been proposed in order to evaluate and improve the success of web sites. Decision-making criteria related to design and content of web sites are needed so that user behavior matches the objectives and expectations of web site owners.

We propose to analyze how well a user's navigation fits company aims, how accurate web site content reflects company's purposes and how much web site structure contributes to achieve company's goals. In order to carry out this proposal, we introduce in this paper an architecture based on software agents. Taking into account that software agents exhibit a degree of autonomous behavior and attempt to act intelligently on behalf of the user for whom they are working [27, 25], web agents included in the architecture deal with all tasks of the proposed method.

In web domain, software agents have been used with several purposes: filtering, retrieval, recommending, categorizing and collaborating. Several systems based on filtering and searching agents have been proposed in order to assist site's users [12, 14, 30, 37].

Based on knowledge represented by multiple ontologies in [28] agents and services are defined in order to support navigation in a conference-schedule domain. ARCH (Adaptive Retrieval based on Concept Hierarchies) [32] helps the user in expressing an effective search query, using domain concept hierarchies.

Most of these systems have been designed as user-side agents to assist users to carry out different kinds of tasks. The agent-based architecture proposed in this paper has been designed taking into account the business point of view. Agents, in our approach could be considered as business-side agents.

The remainder of the paper is organized as follows. Section 2 introduces the way web logs have to be enriched in order to extract the semantics that we need. In section 3 the proposed methodology to calculate relevant factors and consequently estimate the result of the session is further explained. Section 4 describes the agent architecture defined for the deployment of the method analyzed in the previous section. Section 5 presents the experimental results obtained from applying the method. Section 6 presents the conclusions as well as suggestions for further research.

2. Preliminaries of the methodology

Due to different points of view, organizations can define different business goals. For example, the marketing department could be interested in the attractiveness and ease of use of the web site for the user and the department in charge of the design of the pages could be only interested in page design. In contrast, the sales department will be interested in the user buying some products. In order to capture this information, we assume that pages have been enriched with semantic information related both to the content (as already proposed in [6]) and to the business goals. In this paper we estimate, for each goal and for each viewpoint, the result of an ongoing session. In order to do so we assume that we already have information on past sessions and that these sessions have been already classified as success or failure sessions (a set of sessions is chosen as training set and classified by an expert). Then, based on this information we compute the end result of the session by means of a predictive algorithm that estimates the importance of visiting certain pages.
In fact, the algorithm will not only take into account the url of the page itself but the semantics underlying such a page. In order for this to be possible, information on pages has to be enriched with information both related to the semantics of the page, from the point of view of its content, and to the business goals.

Let $Goals = \{g_1, g_2, \ldots, g_s\}$ represent the goals of a company in a particular moment.
Let $Viewpoints = \{v_1, v_2, \ldots, v_n\}$ represent different point of views (e.g., marketing, sales).
Let w_{jj} be the function that assigns weights to each goal, where i represents the i^{th} viewpoint and j the j^{th} goal. We assume that weights are going to be assigned to different goals depending on the viewpoint.
As this function is a weight function, it has the following properties:

- $0 \leq w_{ij} = w(g_i, v_j) \leq 1$ where $g_i \in Goals$ and $v_j \in Viewpoints$
- $\sum_{i \in Goals} w_{ij} = 1$

Table 1 shows an example of possible weight assignment.

	Viewpoints		
Goals	Marketing	Sales	Design
Top of Mind	0.7	0,1	.05
Awards	0.1	0,05	0.9
High sales	0.2	0.85	0.05
Total	1	1	1

Table 1: Possible weight assignment

Let $\Omega = \{\alpha_1, ..., \alpha_m\}$ be the set of pages of a web site. This set includes all pages that can be dynamically generated.

We propose to assign semantics to each page. The approach has no problem with dynamic pages as the semantics is assigned to them when they are generated (it is not related to the url itself but to the content). According to Berendt [4], based on ontologies, an analyst can use several abstract criteria to order pages or groups of pages. In this sense, it is possible to have different ontologies associated to different points of view and business goals. We could also use content-based, service-based hierarchies [7] or domain knowledge contained in the site [13].

In our case, for each site and for each particular viewpoint we define the set of semantics (actions and/or contents) that are relevant to be studied by the site. Notice that this set of semantics can be modified along the site life depending on the things that are relevant to be analysed at each moment. According to this information, ontologies are defined.

Nevertheless, in this paper we are not concerned about how the ontology is created and maintained but about the way we use this information to obtain a predictive model that, using the semantics of the pages, can compute the value of an ongoing session. We will use a granular approach so that the value of a session can be estimated taking into account the semantics of each of the visited pages.

Once the logs have been enriched, sessions are classified as success or failure so that two tables, *Enriched session* and *Results,* are obtained. In order to do so, a training set of sessions is chosen and they are classified according to an expert, so that the result of the session is obtained. The same expert decides the relevant concepts (extracted from the ontologies) to enrich each session.

As a consequence, in the *enriched session* table, each tuple contains, apart from the session identifier, the concepts that are taken into account (i.e., information related to action, contents, design of the pages). For each piece of information taken into account, the session can only take values 0 or 1 to specify whether that action happened during the session or not. As we have already mentioned, this is a granular approach that takes into account relevant actions or behaviours at the site rather than in individual pages.

On the other hand, it is also necessary to obtain a *Results* table that contains information related to the success or failure of each session from each point of view considered. This is innovative in the sense that the proposed algorithm will not only extract relevant concepts for the site but relevant concepts of the site depending on the viewpoint considered. This way, for each session and for each goal we will have a measure of how successful the session happened to be. Once this information is available, the proposed analysis is performed for each viewpoint and for each goal.

3. Proposed Methodology: Predicting an ongoing session success or failure

For each pair (viewpoint, goal), a table is generated in which the attributes contain semantics extracted from the pages that can be visited during a session. The value of each attribute for each session will be 1 if the action took place in that session, or 0 otherwise. The last column of the table specifies, according to an expert, if the session was a success or a failure for a particular (viewpoint, goal) pair. This table is directly obtained from combining the *enriched session* with the *results* tables described above.
In a second step, we applied the stepwise multivariate predictive model to two sets of sessions designated as successful and failure n_1 and n_2 , respectively. The basic strategy in discriminant analysis is to define a linear combination of the dependent attributes in our case representing semantic actions or concepts $s_1, s_2, \ldots, s_t$. So the equation will have the form:

$$L = v_1 s_1 + v_2 s_2 + \ldots + v_t s_t.$$

Once this equation is obtained, the success or failure of a session will be established on the basis of the value of L obtained for that session.

$$If \begin{cases} L \geq 0 \text{ session success is predicted} \\ L < 0 \text{ se session failure is predicted} \end{cases}$$

Discriminant analysis is useful in finding the most relevant semantic concepts. In each step of the stepwise discriminant technique the importance of each attribute included can be studied. This is important not only because we will have the equation to estimate the value of a session in the future but also because the method provides the analyst with a criteria to understand which actions are most relevant for the success of a session from each viewpoint. The computation of the discriminant function has been done according to [18, 2, 1, 24]
The model estimates the coefficients (values) of each attribute considered (semantic) for each pair of goals and viewpoint considered.

From this result it is possible to establish the relationship between semantics and the result of the session. Those coefficients with higher positive value are associated with sessions ending successfully and those taking negative values are associated with failure sessions. So it would be desirable that pages visited by users in their sessions have semantics related to these coefficients associated with them.
So the innovative aspect of the proposed approach has to do with both the predictive model and the measure used to establish concepts or actions that happen during a session to make it successful from a certain site viewpoint. Notice that the number of relevant concepts (*N1*) will be much less than the number of pages *N (N1<<N)* in a web site, so that the problem of analyzing user session decreases in complexity, improving the performance of the methods used to analyze them.

4. Architecture Overview

In order to deploy the methodology described above, it is necessary to define a global architecture, including the modules that deal with the method tasks. The main tasks are:

1. Preprocessing: The result of this task is the Enriched Sessions table.
2. Classification: The result of this task is the Results table.
3. Usage of the predictive model.
4. Refining stage: This task adjusts the model to new results.

Web Mining tasks have been often implemented by agents. The agent paradigm offers desirable features such as autonomy, that is, the ability of acting itself and on behalf of others and proactivity, that is, the ability of acting in anticipation of future problems, needs or changes. These characteristics are very suitable in the scenario described in previous sections because of its dynamic idiosynchrasy.

Thus, a multiagent architecture is proposed, which is composed of three different layers:

1. Semantic Layer: This layer contains agents related to the logic of the algorithm or method used in the ongoing session value estimation.
2. Optimization/Decision Layer: This layer contains agents responsible for optimizing or taking decisions depending on the estimated value.
3. Services Provider Layer: This layer contains agents that provide several services to the rest of the agents. These services are generic and useful tasks, which are independent of the other layers and offer an interface, which may be used by any agent who asks for a service.

Semantic Layer

This level is fed directly by the session value estimation implementation. The agents of this layer deal with the concepts value estimation and its usage in the following sessions. This layer is a multiagent subsystem composed of different specialized agents. They are:

- Preprocessing agents: These agents are responsible for enriching the sessions.
- Classification agents: Although the classification of a session as success or failure is made according to an expert, it is possible to automate this task, through the usage of previous classifications and expert knowledge.
- Estimation agents: These agents apply the stepwise multivariate predictive model. In a first phase, this operation is made offline. Nevertheless, the algorithm must be applied online in current web usage data to estimate the result of a session.
- Refining agents: If the prediction of the used model is not successful, it is necessary to refine the algorithm with new information. This kind of agents must communicate with the estimation agents in order to inform them about the changes.

As we can see, using agents in the semantic layer provides dynamism to the algorithm deployment.

Optimization/Decision-making Layer

This layer includes agents that make decisions or optimize depending on the information supplied by the semantic layer. Although it is possible to build other kind of agents, we have defined the following agents with the aim of optimizing the accesses and personalizing usage of the web:

- Prefetching agents: These agents prefetch most probably next visited web pages, depending on the session value, that is, giving priority to the sessions with a higher value. In this way, the web session load is more efficient and the user feels more comfortable with the web site.
- Adaptive agent: These agents are responsible for building offers adapted to the preferences of the users. These offers may show as popups or web pages. Any other kind of personalization may be added to the logic of these agents.

Services Provider Layer

This level includes generic services used for assisting other layers agents. These agents are:

- Data retrieval agents: These agents goal is to retrieve data from different information sources. These sources are heterogeneous (i.e., databases, files) and they have different types of information and access requirements. Therefore, these agents may delegate on specialized agents for different sources.
- Locator agents: These agents aim to connect and communicate an agent with another one. For finding a given agent, the locator agents use its category, that is, the type of agent. If there are no available agents of this type, the system launches a new agent for serving this request.

It is possible to add more services, implementing the corresponding agent and defining an interface with the rest of the architecture.

Figure 1 represents the three layers of the proposed architecture and their relationships. Notice that there are both internal and external relationships among different kinds of agents. The different information sources are also shown in the graphic.

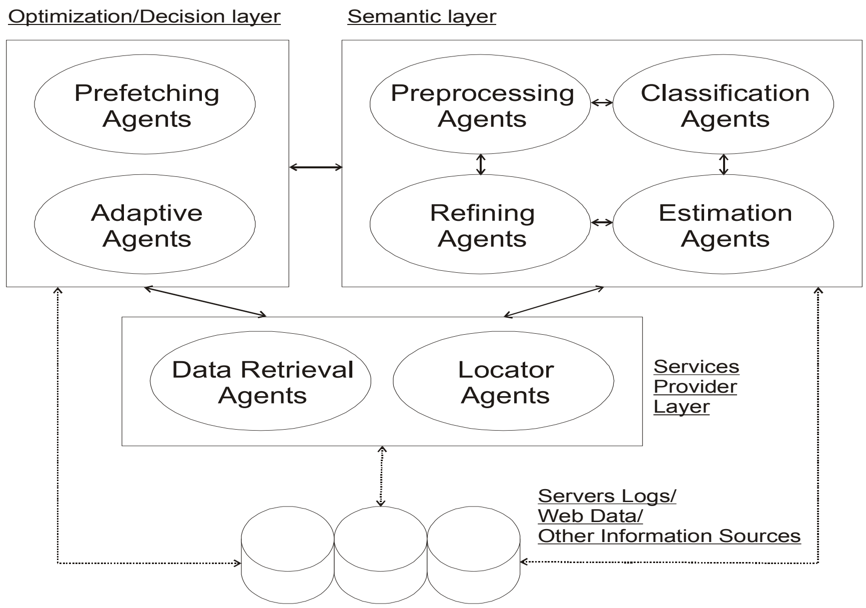

Fig. 1: Web-behavior agent-based architecture layers

5. Case Study - Experimental Results

In order to validate the proposed approach we have applied it to an e-commerce site with 2500 pages. We have preprocessed the log data from accesses to the server during a period of one month. After filtering out irrelevant entries, the data were segmented into 38058 sessions. We considered 20 goals, five points of view and 94 semantic actions. We are dealing with an online shop in which a set of historical sessions has been given that have been classified according to different viewpoints. As an example of these different viewpoints, the success criteria for one department (marketing) is that the user acquires a product while for another department (sales) the success criteria is given by the amount of benefit of the visit (amount sold).
The main goal of the analysis is to estimate the result of a session. Additional benefits of the proposed approach are:

- Establishment of the concepts or actions behind the visited pages (semantics) that are more relevant for the success of a session (i.e., those that contribute most to the site achievement of goals).
- To provide on-line information to automatically make decisions on what motivates the user (e.g., automatic precatching, online offers and discounts).

For each goal and viewpoint of the site, the value of each concept has been established and the results shown in table 2 have been obtained. Notice that it is a good model as it predicts the 93,3% of the successful sessions and the 82,2% of the failure sessions.

	Predicted	
Observed	Success	failure
Success	93,3	6,7
failure	17,8	82,2

Table 2: Summary discriminant table

The results for different viewpoints and goals considered in the experimental results are shown in table 3.

Viewpoint

		1				2			
GOAL		Predicted				Predicted			
1	Observed	Success	Failure	Total	Observed	Success	Failure	Total	Observed
	success	93,3	6,7	100	Success	94,5	5,5	100	success
	Failure	17,8	82,2	100	Failure	14,3	85,7	100	Failure
		Predicted				Predicted			
2	Observed	Success	Failure	Total	Observed	Success	Failure	Total	Observed
	success	85,4	14,6	100	success	87,7	12,3	100	success
	Failure	12,5	87,5	100	Failure	14,6	85,4	100	Failure
		Predicted				Predicted			

Table 3 : Sample of Summary tables

The discriminant function obtained was:

$L=0{,}431\ s_1+,0{,}99\ s_1+\ 0\ ,126\ s_2+0{,}751s_3-0{,}444\ s_4-0{,}107\ s_5+0{,}119\ s_6+0{,}742\ s_7+0{,}306s_8$

The proposed analysis made it possible to establish the relevant factors to take into account when analyzing sessions. Eight relevant concepts to estimate session results were obtained from the 94 initial number of concepts.

In Figure 2, it is shown (Table 2) that most of the sessions were successfully classified. The in-depth analysis of the wrong classified sessions (Figure 3) helped the site sponsor recognize market niches. Note that it is more difficult to describe the wrongly classified examples as they are found in the edges.

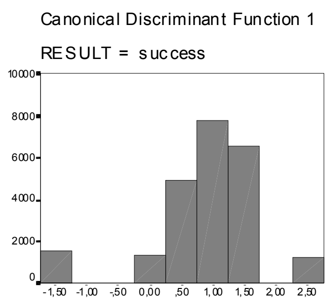

Fig. 2: Canonical Discriminant Function 1 - RESULT = success

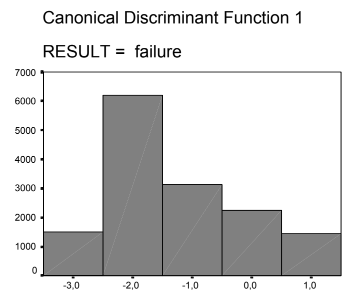

Fig. 3:Canonical Discriminant Function 1 - RESULT = failure

In 83% of the combinations of goals and viewpoints the number of relevant concepts able to classify a session was 20 % or less of the 94 initial concepts.
Finally, when goals were weighted according to each point of view taken into account, an aggregate perception of all goals was obtained for each point of view (i.e., the same session was successful for the Department of Marketing but not for the Sales Department).

6. Conclusions

In this paper we have presented an approach based on discriminant analysis that helps establish semantic factors (e.g., actions, elements) that are relevant for the success of a session. The methodology is also useful for classifying on-going sessions so that actions can be taken to motivate the user to stay longer in the site or to leave it. An innovative aspect of the approach is that the semantics of the pages is established according to different goals and viewpoints so that the analysis can be done according to different criteria. Besides, an agent-based architecture has been defined with the aim of providing dynamism to method deployment. The proposed approach has been applied to an e-commerce site and the results not only helped establish on-line discounts and offers to visitors but also helped the site sponsors discover some market-niches.
New challenges and future research directions address the problem of classifying sessions as success or failure and of having a measure of such success or failure. We are also working on a mechanism to automate this task. On the other hand, we are also analyzing how the sequence of events can be determinant in the success of the session. These open issues could be developed and addressed by multiple alternatives and the forecoming work on this approach will present them.

Acknowledgments

The research has been partially supported by Universidad Politécnica de Madrid under Project WEB-RT and Programa de Desarrollo Tecnológico (Uruguay).

References

1. R. E. Anderson, R. L. Tatham, W. C. Black. Multivariate Data Analysis (5th Edition) J. F. Hair (Editor), Prentice Hall College Div; 5th edition, March 1998
2. T. W. Anderson. An Introduction to Multivariate Statistical Analysis, 2nd Edition Wiley-Interscience; 2 edition, September 1984
3. S. Ansari, R. Kohavi, L. Mason, Z. Zheng. Integrating E-commerce and Data Mining: Architecture and Challenges. In Workshop on Web Mining for E-Commerce - Challenges and Opportunities Working Notes (KDD2000), pages 27-34, Boston, MA, August 2000
4. B. Berent. Detail and Context in Web Usage Mining: Coarsening and Visualizing Sequences, In R. Kohavi, B.M. Masand, M. Spiliopoulou, & J. Srivastava (Eds.), *WEBKDD 2001 - Mining Web Log Data Across All Customer Touch Points* (pp. 1-24). Berlin. Springer, LNAI 2356.
5. B. Berendt. Using Site Semantic to Analyze, Visualize, and Support Navigation. Data Mining and Knowledge Discovery 6, pp.37-59, 2002.
6. B. Berendt, A. Hotho, G. Stumme. Towards Semantic Web Mining. ISWC 2002, LNCS 2342, pp. 264–278, 2002.
7. B. Berendt, M. Spiliopoulou. Analysis of navigation behavior in web sites integrating multiple information systems. The VLDB Journal, 9. pp.56-75, 2000.
8. P. Berthon, L. Pitt, R. Watson. The World Wide Web as an advertising medium. Journal of Advertising Research 36(1) pp.43-54,1996
9. J. Borges, M. Levene. A heuristic to capture longer user web navigation patterns. In Proc. of the First International Conference on Electronic Commerce and Web Technologies, Greenwich, U.K., September 2000.
10. J. Broekstra, A. Kampman, F. Harmelen. Sesame: An Architecture for Storing and Querying RDF Data and Schema Information , Semantics for the WWW, MIT Press, 2001, D. Fensel, J. Hendler, H. Lieberman and W. Wahlster.
11. E. H. Chi, P. Pirolli, J. Pitkow. The Scent of a Site: A System for Analyzing and Predicting Information Scent, Usage, and Usability of a Web Site. In Proc. of CHI '2000 The Hague, Amsterdam CHI Letters volume 2, issue 1.
12. J. Collins, W. Ketter, M. Gini, B. Mobasher. A Multi-Agent Negotiation Testbed for Contracting Tasks with Temporal and Precedence Constraints" Int. Journal of Electronic Commerce, 7(1):35--57, 2002.
13. H. Dai, B. Mobasher. A Road map to More Effective Web Personalization: Integrating Domain Knowledge with Web Usage Mining. Proc.of the International Conference on Internet Computing 2003 (IC'03), Las Vegas, Nevada, June 2003.
14. R. Doorenbos, O. Etzioni, D. Weld. *A Scalable Comparison-Shopping Agent for the World-Wide Web*. In Proceedings of Autonomous Agents (Agents'97), 1997.
15. J. Eighmey. On the Web: It's What You Say and How You Say It. Greenlee School of Journalism and Communication, Iowa State University. [Eighmey Website Response Profile. http://eighmey.jlmc.iastate.edu/DE.html].
16. D. Florescu, A. Levy, A. Mendelzon. Database techniques for the World-Wide Web: A survey. SIGMOD Record (ACM Special Interest Group on Management of Data), 27(3):59.

17. I. Horrocks, S. Tessaris. Querying the Semantic Web: A Formal Approach. ISWC 2002, LNCS 2342, pp. 177–191, 2002.
18. D. G. Keibaun, L. L. Kupper, K. E. Muller Applied Regression Analysis and Other Multivariable Methods. PWS-KENT Publishing Company, USA 1988.
19. A. Magkanaraki, S. Alexaki, V. Christophides, D. Plexousakis. Benchmarking RDF Schemas for the Semantic Web. ISWC 2002, LNCS 2342, pp. 132–146, 2002.
20. E. Menasalvas, S. Millán, E. Hochsztain. A Granular Approach for Analyzing the Degree of Afability of a Web Site. In International Conference on Rough Sets and Current Trends in Computing RSCTC2002. October 2002.
21. B. Mobasher, H. Dai, T. Luo, M. Nakagawa, J. Witshire. Discovery of aggregate usage profiles for web personalization. In Proceedings of the WebKDD Workshop. 2000.
22. Mobasher B., Dai H., Luo T., Sun Y., Zhu J. Integration Web Usage and Content Mining for More Effective Personalization. In Proceedings of the International Conference on E-Commerce and Web Technologies (ECWeb2000), September 2000, Greenwich, UK.
23. B. Mobasher, N. Jain, E. Han, J. Srivastava. Web mining: Pattern discovery from WWW transaction. In Int Conf. on Tools with Artificial Intelligence, pp 558-567, New Port, 1997.
24. D. Morrison. Multivariate Statistical Methods, Duxbury Press; 4th edition. 2002.
25. O. Nasraoiu, R. Krisnapuram, A. Joshi. Mining web access logs using a fuzzy relational clustering algorithm based on a robust estimator, 1998.
26. D. Oberle, B. Berendt, A. Hotho, J. Gonzalez. Conceptual User Tracking. In Advance in Web Intelligence. LNAI 2663. AWIC 2003, Madrid, pp. 155-164, May 2003.
27. M. Papazoglou. Agent-Oriented Technology in Support of e-business. Communications of the ACM, Vol 44, N. 4. 2001.
28. T. R. Payne, R. Singh and K. Sycara. Browsing Schedules - An Agent-Based Approach to Navigating the Semantic Web. I. Horrocks and J. Hendler (Eds.): ISWC 2002, LNCS 2342, pp. 469–473, 2002. Springer-Verlag Berlin Heidelberg 2002.
29. J. Pei, J. Han, B. Mortazavi-Asl, H. Zhu. Mining access patterns efficiently from web logs. In Proceedings Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'00), 2000.
30. B. Sarwar, J. A. Konstan, A. Borchers, J. Herlocker, B. Miller, J. Riedl. *Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System*. CSCII'98 , Seattle WA, 1998.
31. C. Shahabi, A. M. Zarkesh, J. Adibi, V. Shah. Knowledge discovery from user's web-page navigation. In Proceedings of the Seventh International Workshop on Research Issues in Data Engineering, High Performance Database Management for Large-Scale Applications (RIDE'97), pages 20-31, Washington- Brussels - Tokyo, IEEE. 1997.
32. A. Sieg, B. Mobasher, S. Lytinen, R. Burke. Concept Based Query Enhancement in the ARCH Search Agent. Proceedings of the 4th International Conference on Internet Computing (IC'03), Las Vegas, June 2003.
33. M. Spiliopoulou, C. Pohle. Data Mining for Measuring and Improving the Success of Web Sites. Data mining and Knowledge Discovery, (5)1-2, 2001.
34. M. Spiliopoulou, C. Pohle, L. Faulstich. Improving the effectiveness of a web site with web usage mining. In Proceedings WEBKDD99, 1999.
35. J. Srivastava, R. Cooley, M. Deshpande, P. Tan. Web usage mining: Discovery and applications of usage patterns from web data. SIGKDD Explorations, 1:12-23, 2000.
36. Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer, D. Wenke. OntoEdit: Collaborative Ontology Development for the Semantic Web. ISWC 2002, LNCS 2342, pp. 221–235.
37. G. Tewari, P. Maes. Design and Implementation of an Agent-Based Intermediary Infrastructure for Electronic Markets. *EC'00*, October 17-20, 2000, Minneapolis.
38. G. Wolfang, S. Lars. Mining web navigation path fragments. In Workshop on Web Mining for E-Commerce - Challenges and Opportunities Working Notes (KDD2000), pages 105–110, Boston, MA, August 2000.