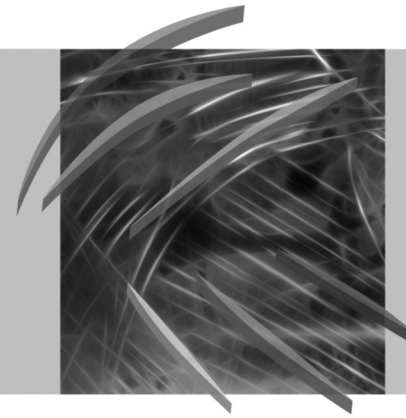
PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu
RESEARCH PAPERS
of Wrocław University of Economics

206

Advanced Information Technologies for Management – AITM 2011

Intelligent Technologies and Applications



edited by
Jerzy Korczak, Helena Dudycz, Mirosław Dyczkowski

Publishing House of Wrocław University of Economics
Wrocław 2011

Reviewers: Frederic Andres, Witold Chmielarz, Jacek Cypryjański, Beata Czarnacka-Chrobot, Bernard F. Kubiak, Halina Kwaśnicka, Antoni Ligęza, Anna Ławrynowicz, Mikołaj Morzy, Stanisław Stanek, Ewa Ziemba

Copy-editing: Agnieszka Flasińska

Layout: Barbara Łopusiewicz

Proof-reading: Marcin Orszulak

Typesetting: Adam Dębski

Cover design: Beata Dębska

This publication is available at www.ibuk.pl

Abstracts of published papers are available in the international database
The Central European Journal of Social Sciences and Humanities http://cejsh.icm.edu.pl
and in The Central and Eastern European Online Library www.ceeol.com

Information on submitting and reviewing papers is available on the Publishing House's website
www.wydawnictwo.ue.wroc.pl

All rights reserved. No part of this book may be reproduced in any form
or in any means without the prior written permission of the Publisher

© Copyright Wrocław University of Economics
Wrocław 2011

ISSN 1899-3192
ISBN 978-83-7695-182-9

The original version: printed
Printing: Printing House TOTEM

Contents

Preface 9

Witold Abramowicz, Jakub Dzikowski, Agata Filipowska, Monika Kaczmarek, Szymon Łazaruk, Towards the Semantic Web's application for preparation of reviews – requirements and architecture for the needs of incentive-based semantic content creation 11

Frederic Andres, Rajkumar Kannan, Collective intelligence in financial knowledge management, Challenges in the information explosion era 22

Edyta Brzychczy, Karol Tajduś, Designing a knowledge base for an advisory system supporting mining works planning in hard coal mines .. 34

Helena Dudycz, Research on usability of visualization in searching economic information in topic maps based application for return on investment indicator 45

Dorota Dżega, Wiesław Pietruszkiewicz, AI-supported management of distributed processes: An investigation of learning process 59

Krzysztof Kania, Knowledge-based system for business-ICT alignment..... 68

Agnieszka Konys, Ontologies supporting the process of selection and evaluation of COTS software components 81

Jerzy Leyk, Frame technology applied in the domain of IT processes job control 96

Anna Ławrynowicz, Planning and scheduling in industrial cluster with combination of expert system and genetic algorithm..... 108

Krzysztof Michalak, Jerzy Korczak, Evolutionary graph mining in suspicious transaction detection 120

Celina M. Olszak, Ewa Ziemba, The determinants of knowledge-based economy development – the fundamental assumptions 130

Mieczysław L. Owoc, Paweł Weichbroth, A framework for Web Usage Mining based on Multi-Agent and Expert System An application to Web Server log files..... 139

Kazimierz Perechuda, Elżbieta Nawrocka, Wojciech Idzikowski, E-organizer as the modern dedicated coaching tool supporting knowledge diffusion in the beauty services sector 152

Witold Rekuć, Leopold Szczurowski, A case for using patterns to identify business processes in a company..... 164

Radosław Rudek, Single-processor scheduling problems with both learning and aging effects 173

Jadwiga Sobieska-Karpińska, Marcin Hernes, Multiattribute functional dependencies in Decision Support Systems 183

Zbigniew Twardowski, Jolanta Wartini-Twardowska, Stanisław Stanek, A Decision Support System based on the DDMCC paradigm for strategic management of capital groups 192
Ewa Ziemba, Celina M. Olszak, The determinants of knowledge-based economy development – ICT use in the Silesian enterprises 204
Paweł Ziemba, Mateusz Piwowarski, Feature selection methods in data mining techniques 213

Streszczenia

Witold Abramowicz, Jakub Dzikowski, Agata Filipowska, Monika Kaczmarek, Szymon Łazaruk, Wykorzystanie mechanizmów sieci semantycznej do przygotowania i publikacji recenzji – wymagania i architektura aplikacji 21
Frederic Andres, Rajkumar Kannan, Inteligencja społeczności w finansowych systemach zarządzania wiedzą: wyzwania w dobie eksplozji informacji..... 33
Edyta Brzychczy, Karol Tajduś, Projektowanie bazy wiedzy na potrzeby systemu doradczego wspomagającego planowanie robót górniczych w kopalniach węgla kamiennego 44
Helena Dudycz, Badanie użyteczności wizualizacji w wyszukiwaniu informacji ekonomicznej w aplikacji mapy pojęć do analizy wskaźnika zwrotu z inwestycji 56
Dorota Dżega, Wiesław Pietruszkiewicz, Wsparcie zarządzania procesami rozproszonymi sztuczną inteligencją: analiza procesu zdalnego nauczania 67
Krzysztof Kania, Oparty na wiedzy system dopasowania biznes-IT 80
Agnieszka Konys, Ontologie wspomagające proces doboru i oceny składników oprogramowania COTS 95
Jerzy Leyk, Technologia ramek zastosowana do sterowania procesami wykonawczymi IT 107
Anna Ławrynowicz, Planowanie i harmonogramowanie w klastrze przemysłowym z kombinacją systemu eksperckiego i algorytmu genetycznego .. 119
Krzysztof Michalak, Jerzy Korczak, Ewolucyjne drążenie grafów w wykrywaniu podejrzanych transakcji..... 129
Celina M. Olszak, Ewa Ziemba, Determinanty rozwoju gospodarki opartej na wiedzy – podstawowe założenia..... 138
Mieczysław L. Owoc, Paweł Weichbroth, Architektura wieloagentowego systemu ekspertowego w analizie użytkowania zasobów internetowych: zastosowanie do plików loga serwera WWW 151

Kazimierz Perechuda, Elżbieta Nawrocka, Wojciech Idzikowski, E-organizer jako nowoczesne narzędzie coachingu dedykowanego wspierającego dyfuzję wiedzy w sektorze usług kosmetycznych 163
Witold Rekuć, Leopold Szczurowski, Przypadek zastosowania wzorców do identyfikacji procesów biznesowych w przedsiębiorstwie 172
Radosław Rudek, Jednoprocesorowe problemy harmonogramowania z efektem uczenia i zużycia 181
Jadwiga Sobieska-Karpińska, Marcin Hernes, Wieloatrybutowe zależności funkcyjne w systemach wspomagania decyzji 191
Zbigniew Twardowski, Jolanta Wartini-Twardowska, Stanisław Stanek, System wspomagania decyzji oparty na paradygmacie DDMCC dla strategicznego zarządzania grupami kapitałowymi..... 203
Ewa Ziemba, Celina M. Olszak, Determinanty rozwoju gospodarki opartej na wiedzy – wykorzystanie ICT w śląskich przedsiębiorstwach 212
Paweł Ziemba, Mateusz Piwowarski, Metody selekcji cech w technikach *data mining* 223

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 206
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS

Advanced Information Technologies for Management – AITM 2011 ISSN 1899-3192
Intelligent Technologies and Applications

Mieczysław L. Owoc*
Wrocław University of Economics, Wrocław, Poland
Paweł Weichbroth*
Memex

A FRAMEWORK FOR WEB USAGE MINING BASED ON MULTI-AGENT AND EXPERT SYSTEM: AN APPLICATION TO WEB SERVER LOG FILES

Abstract: In this paper we introduce a Web Usage Mining (WUM) system, which transparently and dynamically generates content (links) and structure (layout) of the web portal for anonymous visitor. To this purpose we make use of the two software concepts: Multi-Agent and Expert System simultaneously. Multi-Agent system is responsible for discovering knowledge from Web Server log files which requires data preprocessing, data mining, content and structure generating. On the other hand, Expert System supports the software agents by means of cognitive and heuristic knowledge provided and verified by a human expert. The main innovation proposed here is a novel approach that can be used to manage large Web portals.

Keywords: Web Usage Mining, Multi-Agent System, Expert System, Association rule.

1. Introduction and motivation

The 21st century is often called an era of information which is highly available through mass media like press, television, radio and the Internet. The delivered "flood of information" rarely meets preferences and needs of its readers, spectators, listeners and users. Inflexible and static medium is meant to be appropriate for everyone. For the first three information channels, traditional and still the most popular, there is no way to adjust content of broadcasting and published information. Fortunately, the World Wide Web, commonly called Internet, might not have this limitation. In this paper we present a complete software-based solution which can be applied to manage the content and structure of large Web portals. In other words, integration of the Multi-Agent System with Knowledge-based System will be able to deliver flexible and dynamic framework for selecting and matching content and structure to particular user.

Web personalization is [Eirinaki, Vazirgiannis 2003] "the process of customizing a Web site to the needs of specific users", taking advantage of the knowledge

* e-mails: mieczyslaw.owoc@ue.wroc.pl; pawel11739@gmail.com

acquired from the analysis of the web server log files. In [Mulvenna, Anand, Buchner 2000] the goal of the Web personalization is defined as "provide users with the information they want or need, without expecting from them to ask for it explicitly". Nowadays, WUM is seen as a viable technique to discover "hidden" knowledge into Web-related data such as Web Server log files. In particular, it is the process of extracting knowledge form Web users access data by exploiting Data Mining methods and techniques. It can be used for different purposes such as personalization, system improvement and site modification [Baraglia, Silvestri 2004]. Our system fulfills those purposes – moreover the idea of adding expert system to well-known concept of discovering knowledge from web server log files results in gaining more valuable knowledge. Support from domain expert (human) relies on identifying sets (itemsets) which are irrelevant and not reusable for adjusting site content.

The rest of this paper is organized as follows. Section 2 reviews some existing and developing research related to WUM frameworks. Section 3 is a complete description of presented framework in which we distinguish two software-based components: Multi-agent and Expert System. Section 4 is the case study where authors introduce details about the knowledge discovery process and show the results from the first research. In the last section we designate directions of further research, which are intended to verify and validate developing system.

2. Related work

It is evident that Web usage mining might be seen as commercial tool for corporations that invest in the e-business sector. The application of Web usage mining techniques to data gathered from customers' online activity can be successfully implemented to acquire business intelligence by providing high-level knowledge in the form of association rules and patterns that describe consumer navigational and purchasing behavior [Kubiak, Weichbroth 2010]. It is worth pointing out that there exists a large variety of Web analytics' products, ranging from free traffic analysis tools to integrated CRM solutions.

Log analysis is the first stage in Web usage mining and is performed by all the commercially available systems [Eirinaki, Vazirgiannis 2003]. Free software applications are Analog [ANA], Follow 2 [FOL], Webalizer [WBA], WebLog Expert [WEX] and many more. All of them provide the end-user with a set of statistical reports, and some of them like WebLog Expert and Follow 2 track user sessions by delivering specific information about each individual user.

More advanced features are available commercially, which has also built some data mining algorithms along with data warehousing services, reports, diagrams and charts. Some of the most well-know software systems are Elytics [ELY], E.piphany [EPI], NCR [NCR], SAS [SAS]. Another way to provide similar functionality is outsourced services, i.e. IBM Global Services [IBM] or Google Analytics [GOG].

Comparable approach to ours can be found in [Eirinaki, Vazirgiannis 2003], where authors outlined the overall process of usage-based Web personalization in five modules: user profiling, log analysis and Web usage mining, content management, Web site publishing and information acquisition and searching. Each module corresponds to each step of the process.

3. The decomposition of framework

In this section, we give a complete outlook on the system as a whole and every major component respectively. Presented architecture is a consequence of accomplished research on the web server log files. Detailed results are given in [Mikulski, Weichbroth 2009, 2011; Weichbroth 2009a]. The first study reveals the necessity to identify "noise" (artifacts) in collected data. In this case, by defining something as "noise" we have in mind pages and profiles which reveals dependences within local sections (email, dates) or temporal pages (local news available for very short time). Secondly, we discover an intensive movement generated by web crawlers (ants, automatic indexers, bots, Web spiders) – a computer programs that browse the portal in a methodical, automated manner, managing and maintaining by search engines (i.e. Google, Yahoo) to provide up-to-date data, published on the web portal. It motivated us to search an adequacy solution to complement the architecture. Taking into consideration obtained results in the first study [Mikulski, Weichbroth 2009] and required functionality, we decided to extend developing solution with the concept of an Expert System. Therefore, the architecture given in this paper can be viewed as an extension of the Multi-Agent system presented in [Weichbroth 2009b].

In the frame of knowledge discovery process (in the literature on data mining [Fayyad, Piatetsky-Shapiro, Smyth 1996; Fayyad et al. 1996; Mitra, Acharya 2003] also called knowledge discovery in databases, KDD), Figure 1 decomposes nine tasks between Multi-Agent and Expert system. The first one is fully-automated, software-based computer system. whereas the second one is a combination of human expert and knowledge-based with built-on inference engine computer system.

The framework in Figure 1 consists of:

1. **Multi-Agent System** which includes four independent computer programs defined as agents: preprocessing (PA), reasoning (RA), content (CA) and manager (MA).

2. **Expert System** which includes: visualization engine, inference engine and knowledge base.

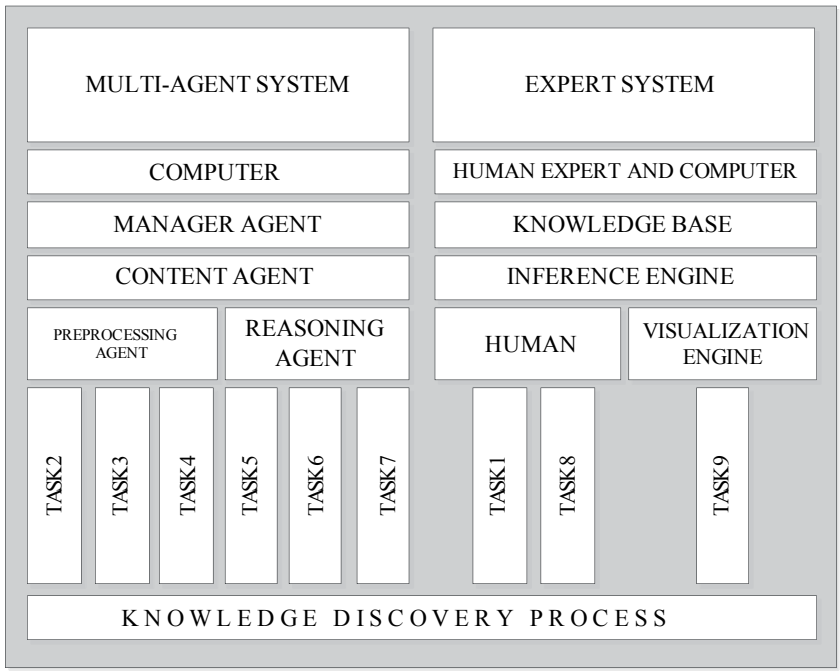


Figure 1. The decomposition of system framework

The knowledge discovery process was divided among nine tasks which were split into Multi-Agent and Expert systems. Principal elements of Web personalization include (a) preprocessing Web server logs, (b) finding frequent itemsets and extracting association rules, (c) identifying and excluding noise from data, (d) grouping association rules to profiles, (e) verifying and validating profiles. All these tasks are briefly described below.

3.1. Multi-Agent System

The review of available literature on multi-agent systems reveals that there are significant differences in a definition of an agent, based on the application of such systems. M. Wooldridge and N.R. Jennings [1995] defined an agent as a computer system located in some environment in which it is capable of autonomous activities according to performed goals.

Software framework where the agents were implemented is JADE (Java Agent Development Framework) [JAD]. The reason we chose this platform is the fact it simplifies the implementation of multi-agent systems through a middleware that complies with the FIPA specifications [FIP] and through a set of graphical tools that supports the debugging and deployment phases.

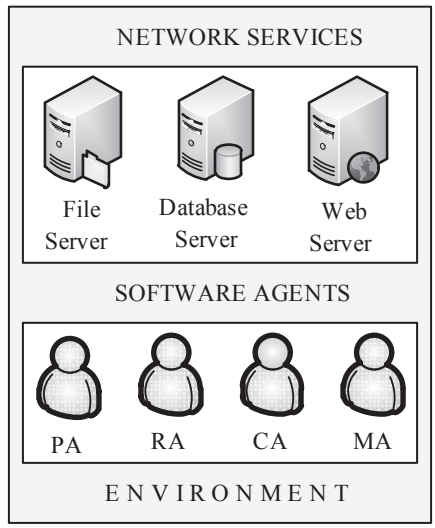


Figure 2. The components of Multi-Agent System

File server is responsible for sharing web server logs files which are intended to be processed by preprocessing agent. We use Mandriva [MAN] Linux distribution as an operating system and Samba [SAM], which together provide secure, stable and fast file service for all clients using the SMB/CIFS protocol. Logically, in a software layer we use an open-source solutions like PostgreSQL [PGR] as a RDBMS (relational database management system), which boasts sophisticated features such as asynchronous replication, Multi-Version Concurrency Control, nested transactions (savepoints), point in time recovery, tablespaces, online/hot backups, a sophisticated query planner/optimizer, and many more. Finally, the www server is Apache [APA] which is the typical one supporting the http protocol [Berners-Lee, Fielding, Frystyk 1995; Fielding et al. 1997, 1999], which source code comes from the National Center for Supercomputing Applications [Team 1996].

Figure 3 shows which tasks are performed by each agent and how they are connected with specific services and resources. We distinguish the following databases:

– WAF (Web Activity Facts): selected data from web server log files,
– WAK (Web Activity Knowledge): frequent itemsets and association rules,
– WAP (Web Activity Profiles): linked rules (based on antecedent and/or consequent) and clusters.

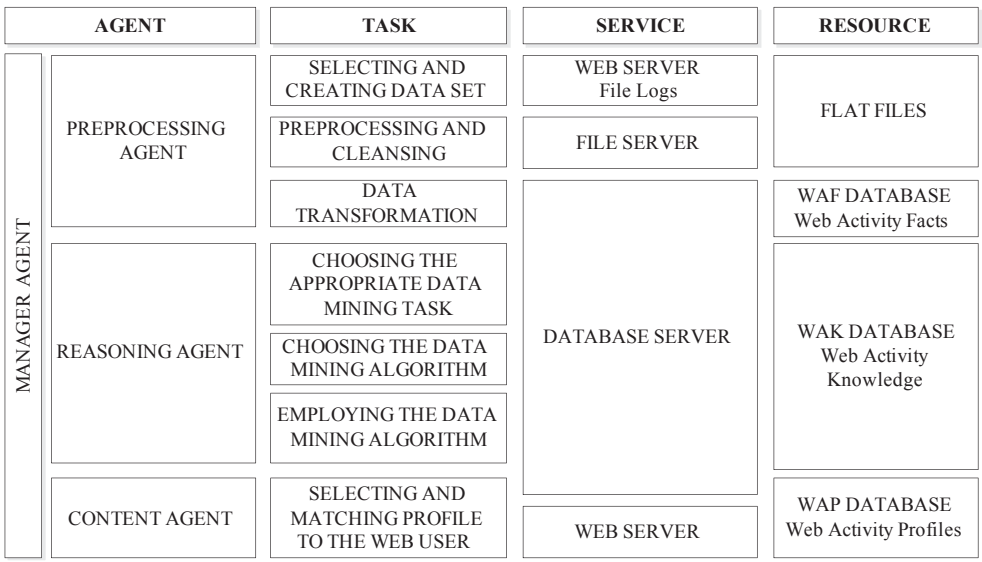


Figure 3. Agents and their particular tasks in connection with services and resources

3.2. Expert System

In [Giarratano, Riley 1998] an expert system was defined as "a computer program that draws on the knowledge of human experts captured in a knowledge base to solve problems that normally require human expertise" and as "a knowledge-based system emulates expert thought to solve significant problems in a particular domain of expertise". Similarly, in [Niederliński 2006] it was defined as "a program which solves problems deputing to experts".

Our expert system is a set of programs to solve problems deputing to experts. It has functional structure which consists of the following components: inference engine, visualization engine, knowledge repository, knowledge base, knowledge base editor, user and expert interface and database server (Figure 4).

Expert system has got all attributes of a typical expert system. The core of the system is inference engine, possessing in two possible inference modes:

– forward chaining,
– backward chaining.

Visualization of the discovered knowledge, depicted in the form of association rules, is processed by visualization engine. Knowledge base is responsible for gathering and sharing knowledge to domain expert. Users' queries and answers are stored in knowledge repository.

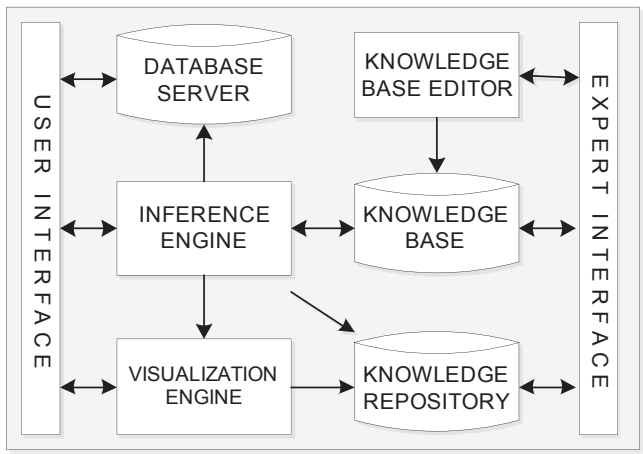


Figure 4. Expert system: functional architecture

COMPONENT	TASK	SERVICE	RESOURCE	
INFERENCE ENGINE	BINDING ASSOCIATION RULES TO PROFILES	DATABASE SERVER	WAP DB	WAK DB
	VALIDATING USERS' PROFILES			
	STATISTICS			
VISUALIZATION ENGINE	VISUALIZATING ASSOCIATION RULES			
	VISUALIZATING USERS' PROFILES			
	PLOTING STATISTICS			
KNOWLEDGE BASE	SELECTING AND MATCHING USERS' PROFILES		KNOWLEDGE BASE DATABASES COPIES : WAF WAK WAP	
	COMPARING USERS' PROFILES			
KNOWLEDGE BASE EDITOR	READING AND MODYFING KNOWLEDGE BASE			
	FORMUALTING NEW KNOWLEDGE			
KNOWLEDGE REPOSITORY	STORING AND MANAGING PROCEDURES		KNOWHOW DATABASE	
	KNOWLEDGE BASE JOBS HISTORY			

Figure 5. Expert system: components and their tasks in connection with services and resources

As can be noticed in Figure 4, in order to edit knowledge base, only expert can do this using knowledge base editor. Inference engine can only have access to the knowledge repository, saving queries declared by users and answers to them, received from inference engine.

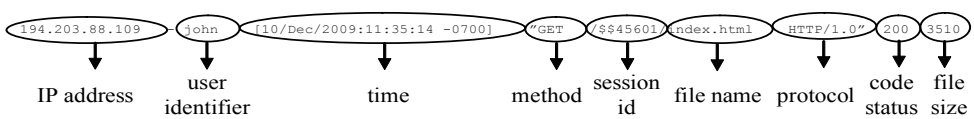
Figure 5 shows which tasks are performed by each component and how they are connected with particular services and resources. In comparison with Section 3, we add knowledge base editor and knowledge repository.

4. Case study

This case study may serve as an example of how the knowledge is extracted from the web server log files. Based on the data which was used in the first research, we will discuss each task of the nine tasks which together can be viewed as the comprehensive web usage mining process.

4.1. Web server log file – an example

Based on one entry which comes from sample web server log file, we examine each variable and later (in next section) focus on those necessary from the research point of view.

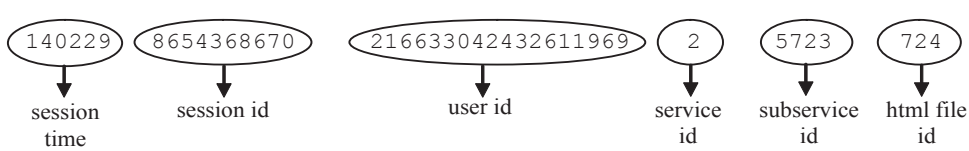


It is worth mentioning that the connection between the client and the WWW server in the http protocol is stateless, so there is no possibility to keep it after the expiry period. Therefore, in order to omit the restriction, the **session identifier** is coded in URL.

4.2. Preprocessing

The preprocessing agent can be seen as a script written in *awk* language combined with *sort* and *uniq* programs used to sort rows of text file and removing multiple copies of the same row.

Data obtained from the portal Onet.pl covers five hours of system activity (from 2PM to 7PM, 29th December 2008). The structure of raw data is presented below where each row of data file has six fields. These data are supplemented with text files containing dictionaries of translations of appearing in the log file identifiers.



The number of numeric variables in the processed log file was reduced from six to two: session identifier (constant variable) and the address of html file path (inconstant variable). In a similar way, the research was conducted on the two other variables: service identifier and subservice identifier. In other words, the discovered knowledge concerns three different variables.

4.3. Discovering knowledge – the users' navigation paths

Data preprocessing was necessary to carry out in order to have data processed properly by reasoning agent. Its main goals are to find all frequents items and generate association rules. The agent was implemented in Java language and thoroughly described in [Mikulski, Weichbroth 2009] where the Apriori algorithm was adopted and modified due to compensate complete agent functionality.

Support ratio was defined arbitrarily on five different levels where its required level is defined every time by researcher. Each element of item represents exactly defined html document (in this case, file extension does not matter) or the name of service or subservice.

Table 1. Results of the first study – frequent items (minimum support 0.01)

Service id	Support (%)	Subservice id	Support (%)	Html file id	Support (%)
Page	78.51	www	78.46	www	80.36
Info	31.64	cnp	27.13	email/cnp/login.html.php3	27.53
Sport	21.06	np	26.58	email/np/dynamic/folder.html	26.29
Business	16.34	world	20.42	email/np/dynamic/folder.html/open.html	15.43
Music	10.37	others	14.83	sport/wc_volleyball/news.html	13.17
Film	08.34	football	14.82	info/world/item.html	12.70
Moto	04.28	market	10.74	info/country/item.html	12.68
Weather	01.19	formula one	9.97	email/np/dynamic/folder.html/delete.html	11.96

Source: own elaboration based on [Mikulski, Weichbroth 2009].

Interpretation of the results given in Table 1 is intuitive – lets us consider first two columns, second row. 31.64% anonymous visitors on 29th December 2008, between 2 PM and 7 PM, have visited information service on the onet.pl portal.

Figure 6 presents the number of frequent itemsets and processing time where the minimum support ratio was set up on 0.01. These results should be considered as the most interesting content for the visitors and should be developed and promoted in the most significant way.

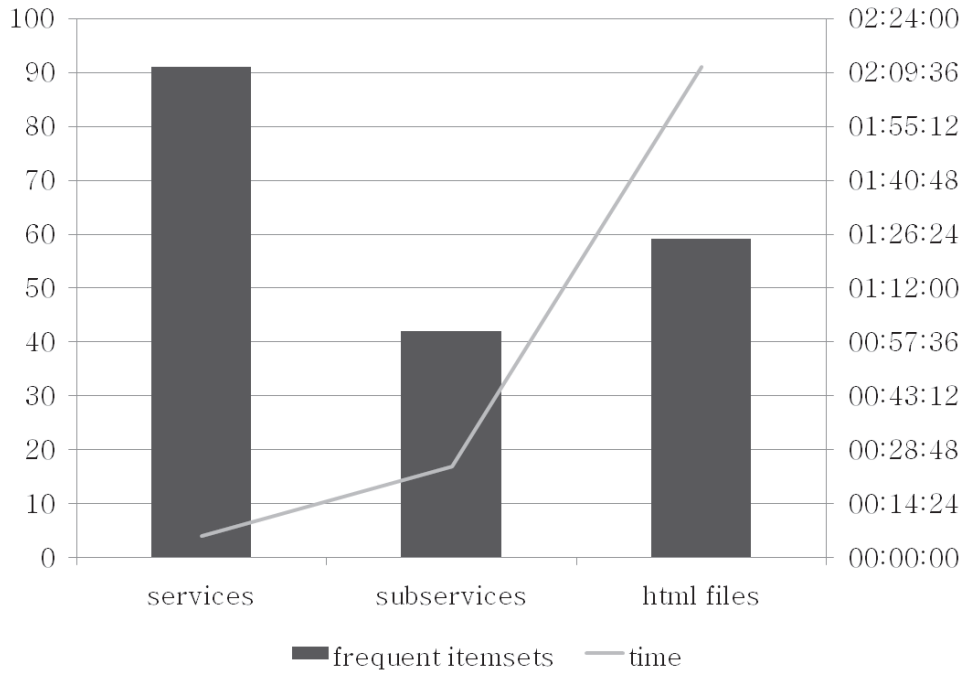


Figure 6. The number of frequent itemsets and processing time

Based on the frequent sets, the reasoning agent generated association rules. Table 2 presents selected association rules built on two, three and four sets, with highest support and confidence ratio where the cut-off was set up on 0.01 and 0.9, respectively.

- *If* **Information** *then* **Page**: 31.6% | 94.2%
- *If* **Sport** *then* **Page**: 21.1% | 94.8%
- *If* **Information** *and* **Sport** *then* **Page**: 10.5% | 96.9%
- *If* **Information** *and* **Business** *then* **Page**: 9.4% | 96.4%
- *If* **Information** *and* **Sport** *and* **Business** *then* **Page**: 3.9% | 97.2%
- *If* **Information** *and* **Sport** *and* **Moto** *then* **Page**: 1.3% | 96.7%

An association rule is a logical statement written in form "if – then". For example, for the last rule, its interpretation would be: 1.3% of users who visited information, sport and moto services, in 96.7% sessions, visited main page also. This rule can be reversed thus can be much more understandable but as a consequence confidence ratio would have different value.

For support on level 1%, based on found frequent items, reasoning agent generated three types of association rules (Figure 7). The number of rules depends on *cardinality of sets* and *confidence ratio* as well.

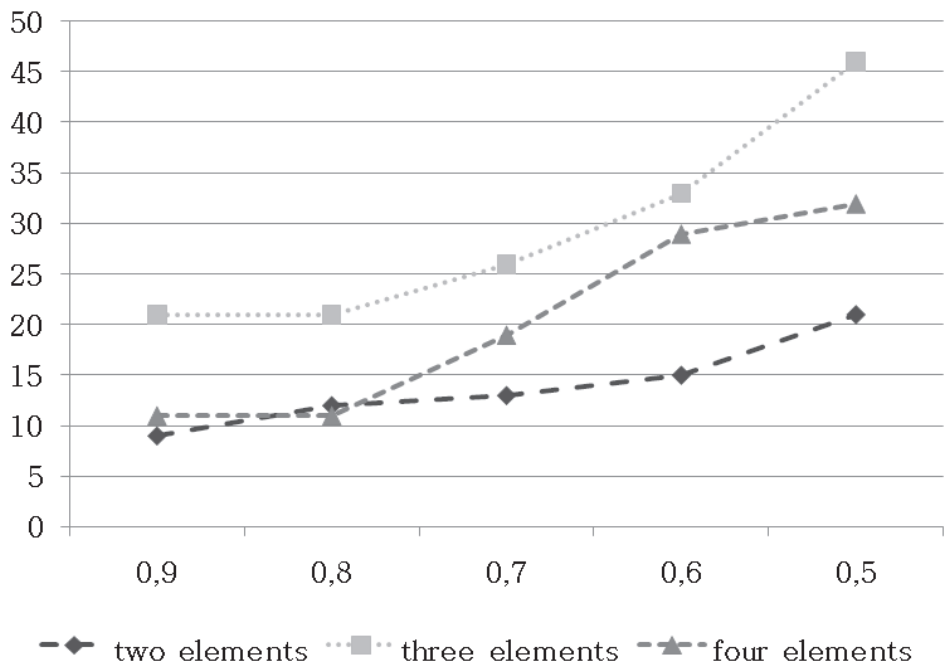


Figure 7. The number association rules based on confidence ratio

The number of variables describes the level of discovered knowledge of users' navigation paths (Figure 7). For assumed minimum support and confidence, it was found that the number of accessed pages in one session did not exceed six (for html file paths). This fact implies kinds of generated association rules.

5. Conclusions

We have briefly described the entire architecture which is intended to discover knowledge from web server logs files and to apply that knowledge to automatically adjust the content and structure of large web portals to needs of its visitors. Moreover, we have shown how knowledge is discovered and briefly present results from the first research.

Presented system will be fully-automated – once executed, operates autonomously and will not require any interaction from the user. Future actions will focus on verifying and validating presented solution on external web portal which, available publicly over the Internet, might point out its weaknesses. On the other hand, implementation and application in real-time mode physical environment will enable to measure its effectiveness and time responses. Presented system is still being de-

veloped and tested – so far the reasoning agent and visualization engine have been implemented.

Appendix A. Web references

[ANA] Analog – http://www.analog.cx.
[APA] Apache server – http://httpd.apache.org.
[ELY] Elytics – http://www.webanalyticsbook.com/.
[EPI] E.piphany – http://www.infor.com/solutions/crm/.
[FIP] FIPA – http://www.fipa.org/.
[FOL] Follow 2 – http://www.mnot.net.
[GOG] Google Analytics – http://www.google.com/analytics/.
[IBM] IBM Global Services – http://www-935.ibm.com/services/us/gbs/consulting/.
[JAD] JADE – http://jade.tilab.com/.
[MAN] Mandriva Linux – http://www.mandriva.com/en/.
[NCR] NCR eCommerce Solutions – http://www.ncr.com/solutions/ecommerce_services.
[PGR] PostgreSQL – http://www.postgresql.org/about/.
[SAM] Samba – http://www.samba.org/.
[SAS] SAS Web Analytics – http://www.sas.com/solutions/webanalytics/.
[WBA] Webalizer – http://www.webalizer.org/.
[WEX] WebLog Expert – http://www.weblogexpert.com/.

References

Baraglia R., Silvestri F. (2004), An online recommender system for large web sites, [in:] *Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, IEEE Computer Society, pp. 199–205.

Berners-Lee T., Fielding R., Frystyk H. (1995), *Hypertext Transfer Protocol – HTTP/1.0. Internet Draft*, http://www.w3.org/Protocols/HTTP/1.0/draft-ietf-http-spec.html (accessed: 20.11.2009).

Eirinaki M., Vazirgiannis M. (2003), Web mining for web personalization, *ACM Transactions on Internet Technologies*, Vol. 3, No. 1, pp. 1–27.

Fayyad U., Piatetsky-Shapiro G., Smyth P. (1996), From data mining to knowledge discovery in databases, *AI Magazine*, Vol. 17, No. 3, http://www.aaai.org/ojs/index.php/aimagazine/article, pp. 37–54.

Fayyad U., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (1996), *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, Cambridge, MA.

Fielding R., Gettys J., Mogul J., Frystyk H., Berners-Lee T. (1997), *Hypertext Transfer Protocol – HTTP/1.1. Internet Official Protocol Standards (RFC 2068)*, http://tools.ietf.org/html/rfc2068 (accessed: 20.11.2011).

Fielding R., Gettys J., Mogul J., Frystyk H., Masinter L., Leach P., Berners-Lee T. (1999), *Hypertext Transfer Protocol – HTTP/1.1. Internet Official Protocol Standards (RFC 2616)*, http://www.w3.org/Protocols/rfc2616/rfc2616.html (accessed: 30.11.2009).

Giarratano J., Riley G. (1998), *Expert Systems Principles and Programming*, PWS Publishing Company, Boston, MA.

Kubiak B.F., Weichbroth P. (2010), Cross- and up-selling techniques in e-commerce activities, *eCommerce, ePayments and New Entrepreneurship*, Vol. 15, No. 3.

Mikulski Ł., Weichbroth P. (2009), Discovering patterns of visits on the Internet web sites in the perspective of associative models, *Polish Journal of Environmental Studies*, Vol. 18 (3B), pp. 267–271.

Mikulski Ł., Weichbroth P. (2011), Usuwanie artefaktów w wykrywaniu wzorców użytkowania stron WWW, [in:] H. Dudycz, J. Korczak (Eds.), *Informatyka w biznesie*, Prace Naukowe Uniwersytetu Ekonomicznego nr 212, Informatyka Ekonomiczna 22, Wydawnictw Uniwersytetu Ekonomicznego, Wrocław.

Mitra S., Acharya T. (2003), *Data Mining. Multimedia, Soft Computing and Bioinformatics*, John Wiley & Sons, Hoboken.

Mulvenna M.D., Anand S.S., Buchner A.G. (2000), Personalization on the net using web mining, *Communications of the ACM*, Vol. 43, No. 8, pp. 123–125.

Niederliński A. (2006), *Regułowo-modelowe systemy ekspertowe rmse*, Wydawnictwo Pracowni Komputerowej Jacka Skalmierskiego, Gliwice.

Team N. H. D. NCSA HTTPd, http://hoohoo.ncsa.illinois.edu/ (accessed: 20.11.2009).

Weichbroth P. (2009a), Analiza zachowań użytkowników portalu onet.pl w ujęciu reguł asocjacyjnych, [in:] A. Grzech, K. Juszczyszyn, H. Kwaśnicka, N.T. Nguyen (Eds.), *Inżynieria wiedzy i systemy ekspertowe*, Akademicka Oficyna Wydawnicza Exit, Warszawa, pp. 81–88.

Weichbroth P. (2009b), The system framework for profiling the content of web portals, [in:] J. Korczak (Ed.), *Data mining and Business Intelligence*, Research Papers of Wrocław University of Economics No. 104, Business Informatics 16, Publishing House of Wrocław University of Economics, Wrocław, pp. 186–193.

Wooldridge M., Jennings N.R. (1995), Intelligent agents: Theory and practice, *The Knowledge Engineering Review*, Vol. 10, No. 2, pp. 115–152.

ARCHITEKTURA WIELOAGENTOWEGO SYSTEMU EKSPERTOWEGO W ANALIZIE UŻYTKOWANIA ZASOBÓW INTERNETOWYCH: ZASTOSOWANIE DO PLIKÓW LOGA SERWERA WWW

Streszczenie: W niniejszym artykule przedstawiono architekturę systemu do analizy użytkowania zasobów internetowych, udostępnianych w ramach usługi WWW. Funkcjonalność systemu zakłada, iż treść oraz struktura portalu internetowego jest dynamicznie i transparentnie generowana dla anonimowego użytkownika. W tym celu wykorzystano dwie znane koncepcje sztucznej inteligencji: systemu wieloagentowego oraz ekspertowego. System wieloagentowy jest odpowiedzialny za ekstrakcję wiedzy z plików loga serwera WWW, co wymaga przetwarzania wstępnego oraz drążenia danych. System ekspertowy wspiera działanie systemu wieloagentowego na poziomie oceny i walidacji odkrytej wiedzy poprzez bazę wiedzy, którą dostarcza i weryfikuje ekspert domeny (człowiek). Wkładem własnym autorów jest włączenie koncepcji systemu ekspertowego, złożonego z maszyny wnioskującej, maszyny wizualizacji oraz bazy wiedzy, do znanej w literaturze przedmiotu koncepcji programowego systemu wieloagentowego, służącego analizie logów serwera WWW.

Słowa kluczowe: użytkowanie zasobów internetowych, system wieloagentowy, system ekspertowy, reguła asocjacji.