User Identification in the Process of Web Usage Data Preprocessing

Jozef Kapusta (✉)
The Philosopher University in Nitra, Nitra, Slovakia
Pedagogical University of Cracow, Cracow, Poland
`jkapusta@ukf.sk, jkapusta@up.krakow.pl`

Michal Munk, Dominik Halvoník, Martin Drlík
The Philosopher University in Nitra, Nitra, Slovakia

Abstract—If we are talking about user behavior analytics, we have to understand what the main source of valuable information is. One of these sources is definitely a web server. There are multiple places where we can extract the necessary data. The most common ways are to search for these data in access log, error log, custom log files of web server, proxy server log file, web browser log, browser cookies etc. A web server log is in its default form known as a Common Log File (W3C, 1995) and keeps information about IP address; date and time of visit; accessed and referenced resource. There are standardized methodologies which contain several steps leading to extract new knowledge from provided data. Usually, the first step is in each one of them to identify users, users' sessions, page views, and clickstreams. This process is called preprocessing. Main goal of this stage is to receive unprocessed web server log file as input and after processing outputs meaningful representations which can be used in next phase. In this paper, we describe in detail user session identification which can be considered as most important part of data pre-processing. Our paper aims to compare the user/session identification using the STT with the identification of user/session using cookies. This comparison was performed concerning the quality of the sequential rules generated, i.e., a comparison was made regarding generation useful, trivial and inexplicable rules.

Keywords—Web usage mining, cookies, session time thresholds, sequence rules

1 Introduction

Public or even private websites are often under various analyzes. The purpose of this analyzes multiple, from maintenance planning to structural improvements. All of this analyzes are based on data. This data can be found in specific places. One of the most valuable sources is a web server. These machines contain a large number of different data sets. For our purposes, we will focus on data related to visitors' behaviour. While the primary web data is used to get knowledge from the web struc-

ture or its content, the secondary one is used in the web usage mining (WUM) [1]. These secondary data sources are usually files placed in the web server like access log, error log customized application log files, or files on other related communication components like proxy server log file, web browser log, browser cookies etc. A web server log is in its default form known as a Common Log File (W3C, 1995) and keeps information about the IP address; date and time of visit; accessed and referenced resource. If we use the extended version of the log file, we can collect more information, e.g. type of browser (User-Agent Field).

To achieve relevant results, it's crucial to follow all standardized steps defined in selected data mining methodology. It's part of the best practice that we always start with data pre-processing. Based on logged data we can identify users, sessions, page views, and clickstreams. The main goal of this stage is to receive unprocessed web server log file as input and after processing, outputs meaningful representations which can be used in next phase. One of the crucial partial steps in data pre-processing is the usage of data cleaning methods. The reason why this step is so important is that web usage mining is sensitive to noise. Based on the above, data pre-processing can be can be a challenging step, especially if the data provided are incomplete or even incorrect. This can cause misleading the results of the entire analysis. According to Cooley, Mobasher and Srivastava [2] data pre-processing consists of :

- Cleaning (adding missing references due to caching mechanisms, removing erroneous and irrelevant references, etc.)
- Integration (integration of registration data, synchronizing data from related server logs, etc.)
- Transformation (user-session identification [3], path completion [4], etc.)
- Reduction (reducing dimensionality) [5]

We discussed in detail the way how in WUM is pattern discovery made in [6]. Pattern discovery is the main outcome of the proposed methodology.

Our main goal is to describe in detail user session identification, which is the third step of data pre-processing according to Cooley, Mobasher and Srivastava. The best approach in this manner will be to use cookies technology for user identification for WUM. This method will provide most relevant data in terms of visitors' behavior. Cookies can contain various information about user, including the session identifier or specific customized data like time spend on web page or other additional data which are usually calculated based on the logged data in later stages of pre-processing stage. The Web server can receive this information every time when user visits a web page. This means that they can be stored in a log file along with other additional data like URL address of requested web page.

Individual visitors can be differentiated also based on the identification of sessions. Intention of session identification is to merge multiple rows in log files representing individual access to separate relations. The definition of these relation can be done in various ways. More prominent strategies define the session as a series of steps that lead to the achievement of a certain goal or that lead to achieve a certain task. One with less demanding methods, Session Time Thresholds (STT), sets a static time win-

dow, e.g. 30 minutes, and considers a one session records logged inside defined time window.

Our paper aims to compare these two described user/session identification methods, where for the session identification we will use STT. The comparison of the user/session identification will adapt following levels of data pre-processing (identification using cookies with and without the path completion vs identification using STT with and without the path completion). In this paper we will not focus on empirically prove that the user/session identification using cookies gives more accurate results according to the quality of extracted sequence rules. We aim to identify the precision of the STT method compared with the identification based on cookies. Writing a new document with this template

You may also simply delete all the text in this document, paste yours and format it with the styles.

2 Literature Review

There are some problems that disallow cookies usage in data-preprocessing. The most widespread are the absence of permission to store them in browsers or even not to support this functionality at all [7]. It's not unusual that browsers automatically clean they cache memory after some time, which also deletes all the cookies information. There is also a scenario when users delete their cookies manually. This will cause the server to fail to recognize the returning visitors [8]. It is necessary to say that cookies are the most widely known and easy-to-use method used for user identification with high usability and level of security.

It is possible to use time thresholds to expose user session limits [9]. The STT method using standard time threshold (time-window) represents the most common technique. Using this method, each time we had found subsequent records about the web page requests where the time of the web page displaying had been higher than explicitly selected time, we divided the user visits into several sessions. This method is popular for its simplicity to implement, and it is often used with various values of STT. The correct window length has a direct impact on the quality level and amount of discovered behaviour patterns. It is customary to set STT values for 5 minutes [10], 15 minutes and 30 minutes [11]. There are several studies that directly examined [12] the suitability of STT values of 15, 30 and 60 minutes, and then compared output from quality perspective. In most of these papers were used either in combination with or without path completion. The results have shown that the use of lower STT values (15 and 30 minutes) has a direct impact on decreasing the amount of trivial and inexplicable rules.

It has to be said that it is possible to use also unorthodox STT values such as 25.5 min in [13]. This value was been obtained as the sum of the average time spent on the web page, what was in this case 9.3 minutes, and it was also taken to account 150% of the standard deviation of the time which user spent on the web page. As a result of these experiments, the STT value of 30 minutes is now considered the most appropriate [14] even if it is not always optimal. Proof of this claim was provided by Toan

Huynh and James Miller. They realized experiments with dynamic STT. The value of the STT was dependent on type of web service [15]. Zhang and Ghorbani [16], also dealt with similar issues, but the results of their experiments can be considered better. Their experiment was related to comparisons of 10 minutes, 30 minutes and dynamic STT.

3   Methodology

The following steps need to be taken when examining the impact of individual steps of data pre-processing on the quality of knowledge:

1. Data acquisition – defining the observed variables into the log file with regard to obtaining the necessary data (IP address, date and time of access, URL address, etc.)
2. Creation of data matrix – from the log files (access data) and the site map (content data)
3. Data preparation on different levels of log file pre-processing
4. Data analysis – discovering user behavioral patterns. While searching user behavioral patterns in the surveyed samples, it is necessary to ensure that the rules from different files are extracted under the same conditions. We use module Sequence, Association and Link Analysis in the system STATISTICA for the extraction of sequence rules. It is an implementation of Apriori algorithm [17-19].
5. Understanding data after the analysis

We create data file based on the output of the analysis and calculate basic characteristics of examined files:

- Accesses count
- Number of identified sequences
- Number of frequent sequences
- Average length of identified sessions

Subsequently, it is possible to specify conditions based on these initial results.

6. The comparison of acquired knowledge from the examined files (pre-processed with different levels of data preparation).

When evaluating the acquired knowledge, we focus not only on the quantity of extracted rules, but also on their quality. Quality of sequence rules is assessed by two attributes [20]:

- Support
- Confidence

When evaluating the acquired knowledge, we focus not only on the quantity of the extracted rules, but also on their quality. We evaluate the acquired knowledge in terms of quantity and quality of the sequence rules found by following these steps:

- Comparison of the percentage of the rules found in the examined files
- Comparison of the support and confidence rules found in the examined files

The very nature of cookies technology makes it obvious that this technology provides the best user identification for WUM. Cookies can contain a wealth of information about users, including the session identifier. This information can be sent back to the Web server whenever the user visits a web page. Thereafter it may be stored in a log file along with a record of the page requests.

Cookies are often restricted in web browsers so their usage in data pre-processing is often impossible. For that reason, we use the heuristic session identification technique using the session time threshold - STT. The aim of our experiment is to compare the user/session identification using cookies with the identification of user/session using the STT. We will compare the user/session identification according to following levels of data pre-processing (identification using cookies with and without the path completion vs. identification using STT with and without the path completion). It is not necessary to experimentally prove that user/session identification using cookies gives better results according to the quantity and quality of extracted sequence rules. Our aim is to determine the reliability of the user/session identification method using the STT compared with the identification based on cookies.
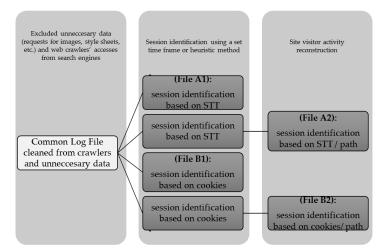

Fig. 1. Experiment research design

4 Experimental Data Preparation

For our experiment research we used the log file of the university website from Constantine the Philosopher University in Nitra. Before we started we had had to allow reading and writing of cookies at the university website and reconfigure the webserver to also log cookies into the log file. After the period of data gathering ended, we cleaned the log file from all unnecessary data (request for pictures, style sheets

etc.) and crawlers requests. The cleaned log file contained 412360 records for the examined period of three weeks. Data pre-processing phase consists of the following levels:

- AY – data with the user/session identification using the STT
- BY – data with user/session identification using cookies
- X1 – without the path completion
- X2 – with the path completion

  Where X = {A, B} and Y = {1,2}
  We can get different files using some combination of the above mentioned levels:

- (**File A1**): user identification using the STT without path completion,
- (**File A2**): user identification using the STT with path completion,
- (**File B1**): user identification using cookies without path completion,
- (**File B2**): user identification using cookies with path completion.

We used the time window of 10 minutes in both cases of user/session identification. We set cookies lifetime to 10 minutes to be able to compare the results.

5   Experiment Results

While analyzing user/session identification using cookies, we identified 2377 (Table 1) more sequences while keeping the same number of records. We can say that the number of sequences increased by 3.5%. On the contrary, the number of frequent sequences decreased by 8.5 %. After path completion in user/session identification using STT, the number of records increased by 23% and the average length of sequences changed from 6 to 7. Higher growth was in the file with user/session identification using cookies. The number of records increased by 39% and the average length of session changed to 8.

Table 1. Number of sequences in each file

	Number of records	Number of visits - identified sequences	Average length of identified sequences	Frequent sequences
File A1	412360	68773	6	111
File A2	508729	68773	7	180
File B1	412360	71150	6	102
File B2	573931	71150	8	184

We stated following assumptions based on the first results:

- We assume that the user/session identification using cookies doesn't influence the quantity of extracted rules
- We assume that path completion will have a significant impact on the quantity and quality of extracted rules

- We assume that user/session identification using cookies in comparison with STT will have a significant impact on the quantity and quality of extracted rules, when using files after path completion

We examined accesses of users of the university website during three weeks in this experiment. The result of the analysis (Table 2) of gathered data is sequence data obtained from frequent sequences that fulfil the minimal value of variable support (in our case min s = 0.005). Frequent sequences were chosen from all identified sequences, i.e. accesses of users on the portal during examined period.

Table 2. Discovered sequence rules in each file

Body	=>	Head	A1	A2	B1	B2
(/), (/ admissions)	=>	(/admissions / admissions-results)	1	1	1	1
⋮	=>	⋮	⋮	⋮	⋮	⋮
(/university-structure)	=>	(/ university-structure)	0	1	0	1
(/university-structure)	=>	(university-structure /faculty-of-natural-sciences)	1	1	1	1
⋮	=>	⋮	⋮	⋮	⋮	⋮
(/study/accredited-study-programmes)	=>	(/study)	0	1	0	1
Count of derived sequence rules			51	197	43	227
Percent of derived sequence rules (Percent 1's)			21.52	83.12	18.14	95.78
Percent 0's			78.48	16.88	81.86	4.22
Cochran Q test			Q = 443.3120, df = 3, p < 0.000000			

There is a high similarity among the results of sequence rule analysis (Table 2) regarding the percentage of discovered rules in files without path completion (A1, B1), as well as in the files with path completion (A2, B2). We extracted most rules from the files with identified users/sessions with path completion. Specifically, 227 rules were extracted from file B2, which means more than 95% and 197 from file A2, which means more than 83% of all discovered rules. Generally speaking, more rules were discovered in files after path completion (A2, B2).

The results of Cochran Q-test are visualized on Figure 2. Based on the results of the Cochran Q test (Table 2) the zero hypothesis is rejected at the 1% significance level. The null hypothesis claims that the occurrence of rules does not depend on the level of WUM data preparation. The low value of Kendal's coefficient (nearly 0.62 (Table 3)), which represents the degree of concordance in the number of discovered rules (value of one represents perfect match and the value of zero represents total difference), proves the result of the Cochran Q-test.
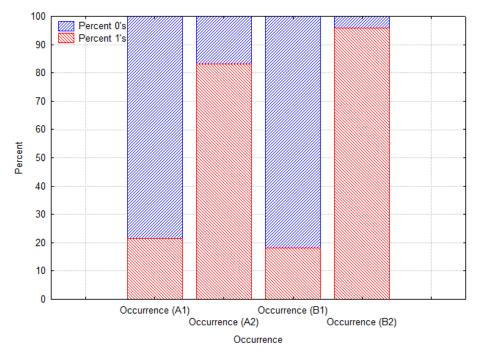
Fig. 2. Graph representing discovered rules on examined files

Table 3. Discovered sequence rules in each file

Occurrence	Mean	1	2	3
File (B1)	0,181435	****		
File (A1)	0,215190	****		
File (A2)	0,831224		****	
File (B2)	0,957806			****
Kendall Coeff. of Concordance				0,62350

We identified one homogenous group based on the multiple comparisons (Tukey HSD test, Table 3). This group consists of files A1 and B1 relative to the direct occurrence of discovered rules. There are statistically significant differences with the notable level of 0.05 in average occurrence of discovered rules among the file A2 and all other files; and among file B2 and all other files; and between file A2 and B2.

Path completion has significant impact on the quantity of extracted rules. The identification of user/session using cookies does not have significant impact on the quantity of extracted rules when we compare it to the identification with STT and without path completion (A1, B1). On the other hand, we prove that path completion has significant impact on the quantity of extracted rules (A2 vs. A1, B1), (B2 vs. A1, B1). Identification of user/session using cookies has significant impact on the quantity of extracted rules when compared to identification using STT only when we complete paths after the identification (A2 vs. B2).

The quality of sequence rules can be evaluated using two indicators - support and confidence. Differences in the results of sequence rule analysis are not only in the quantity of discovered rules, but also in the quality (the value of support variable) of discovered rules in examined files. Kendall's coefficient of concordance represents the level of conformity in support of discovered rules.

Table 4. Homogenous groups for support characteristic of discovered rules

Support	Mean	1	2
File (B1)	1.492270	****	
File (A1)	1.565474	****	
File (A2)	1.749596		****
File (B2)	1.781953		****
Kendall Coeff. of Concordance			0.25132

Table 5. Homogenous groups for confidence characteristic of discovered rules

Confidence	Mean	1	2
File (B1)	24.53741	****	
File (A1)	24.62301	****	
File (A2)	26.70755		****
File (B2)	27.01388		****
Kendall Coeff. of Concordance			0.02880

We identified three homogenous groups using multiple comparisons (Tukey HSD test) in both cases. We showed statistically significant differences (Table 4, Table 5) among files with path completion (A2, B2) and files without path completion (A1, B1) with the significance level of 0.05 in support and confidence of discovered rules.

When we focus on the difference between the quantity of discovered rules (Table 3) and the quality of discovered rules (Table 4, Table 5), we can see the difference only in files with path completion. From the perspective of the quality, files A2 and B2 belongs to one group. It means that from the perspective of the quality of discovered rules there is no statistical significant difference between identification of users/sessions using STT and using cookies. Statistically significant differences are present only when we use path completion.

6 Discussion

Statistically we compared the methods of user/session identification using the cookies technology and STT in the experiment. In the entire experiment, deliberately we use the phrase 'user/session identification". Cookies identify the user, but in the case of setting the expiration of created cookies (in our case 10 min.) it can be also used for identifying sessions. Identifying of user IP address in the case of STT in the time-oriented heuristics for identifying sessions is essential. This is insufficient to identify users. STT method first of all identifies sessions, and of course at a suitable threshold time also often identifies the users.

The aim of the experiment was to compare the user/sessions on four different levels of data preparation and also determine how reliable of a method it is to identify users/sessions by STT.

Most of the conditions for data collection "discriminated" method STT and were more suited for cookies method. Among the positive factors in favor of cookies method resulting from the analyzed files from university website, we can include the following:

- For the analysis the period during the semester (end of winter semester 2016/2017) was selected, when it is likely that students increasingly use the portal of university computer labs or internal.
- As part of the university (computer students in the hostels, accesses the UKF network via WiFi and computers in classrooms) access to the internet through NAT gate, i.e. most computers on the network UKF are hidden as a small number of IP addresses in the log file.
- A computer's mechanism for "cleaning the students' activity" is implemented in most university classrooms. It renews the default settings of the data stored on the computer after signing out; respectively signing in new student (each student has their own user account). This means the old cookies always get deleted despite the possible settings for cookies expiration.

Despite these positive conditions in favor of the cookies, the described experiment did not show a statistically significant difference between the identification of users/sessions via cookies and the identification using the STT in terms of the quality of the extracted rules. The only proven statistically significant difference was demonstrated in the quantity of the extracted rules. This was demonstrated when path completion was applied. A statistically significant difference between exploring methods has not been demonstrated in terms of the quantity if we did not apply the path completion algorithm to the source data.

For these reasons, we can conclude that the method is more suitable for STT users/sessions identification in terms of quality of the extracted rules. From the obtained rule quality point of view the STT's suitability is ambiguous in comparison with cookies.

We observed a major drawback in cookies usage, which are also demonstrated in our experiment, we have noticed limitations of this technology. In our experiment, we failed to record cookies for 28.04% of all log file records. Mainly for this reason, we consider it a more appropriate time-oriented techniques for identifying a user / session.

7 Conclusion

A prerequisite of good analysis is reliable data. The raw data from the web server log file contains unnecessary, irrelevant, inaccurate and incomplete information about using the web. Unnecessary data creates lines of log file in which there are recorded requests for images, styles and scripts and other files that can be embedded in the

page. Irrelevant data represents lines of log file in which accesses are recorded not by users - site visitors whose behavior is analyzed, but by robots of various search services. The inaccuracy of the data relates to the anonymous nature of the data. We consider that the log file is a resource of anonymous data about the user because we do not record their personal data or data that can be used to uniquely identify the user. Incompleteness of data causes the browser cache. The log file records contain lacks of paths, that user enters through the "Back" button.

For a comprehensive view on the issues of data preparation for WUM, in this paper we realized new experiments to prepare data using other heuristic methods. The key technology for session identification based on time was compared with the unique identification technologies using cookies. The use of unambiguous identification technology using cookies in practice faces several problems due to the fact that this method is not one of the widely used. These problems were firstly mentioned in the beginning of the experiment. In our first experiment, 28% of all records did not set records of cookie IDs in the log file because of technical issues.

A prerequisite for any web usage data analysis, no matter for what purpose it is applied (traffic analysis, restructuring, optimization of website, etc.) is the data itself. Results of the analysis depend on the quality of the analyzed data. On the other hand, data preparation is a time-consuming stage of the knowledge discovery process. The results of our experiments are for each portal that is regularly reviewed and adapted very important because it shows the correctness of each step in the analysis or more precisely, identification of "unnecessary" steps alone can simplify data preparation. If we see a technological problem, time, and especially the money for its implementation, for every step of data preparation, it is clear that the results of experiments are very important and necessary for practice.

8   References

[1] B. Liu, Web data mining, New York: Springer, 2007.

[2] R. Cooley, B. Mobasher, and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns," Knowledge and Information System, vol. 1, 1999. https://doi.org/10.1007/BF03325089

[3] Z. Chen, A. W.-C. Fu, and F. C.-H. Tong, "Optimal Algorithms for Finding User Access Sessions from Very Large Web Logs," Advances in Knowledge Discovery and Data Mining: 6th Pacific-Asia Conference, PAKDD 2002 Taipei, Taiwan, May 6-8, 2002 Proceedings, M.-S. Chen, P. S. Yu and B. Liu, eds., pp. 290-296, Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. https://doi.org/10.1007/3-540-47887-6_28

[4] Y. Li, B. Feng, and Q. Mao, "Research on Path Completion Technique in Web Usage Mining," in Proceedings of the 2008 International Symposium on Computer Science and Computational Technology - Volume 01, 2008, pp. 554-559. https://doi.org/10.1109/ISCSCT.2008.151

[5] O. Nasraoui, and E. Saka, "Web Usage Mining in Noisy and Ambiguous Environments: Exploring the Role of Concept Hierarchies, Compression, and Robust User Profiles," From Web to Social Web: Discovering and Deploying User and Content Profiles: Workshop on Web Mining, WebMine 2006, Berlin, Germany, September 18, 2006. Revised Selected

and Invited Papers, B. Berendt, A. Hotho, D. Mladenic et al., eds., pp. 82-101, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. https://doi.org/10.1007/978-3-540-74951-6_5

[6] S. P. Nina, M. Rahman, K. I. Bhuiyan et al., "Pattern discovery of web usage mining." pp. 499-503.

[7] J. Srivastava, R. Cooley, M. Deshpande et al., "Web usage mining: discovery and applications of usage patterns from Web data," SIGKDD Explor. News, vol. 1, no. 2, pp. 12-23, 2000. https://doi.org/10.1145/846183.846188

[8] A. Guerbas, O. Addam, O. Zaarour et al., "Effective web log mining and online navigational pattern prediction," Knowledge-Based Systems, vol. 49, no. 0, pp. 50-62, 2013. https://doi.org/10.1016/j.knosys.2013.04.014

[9] D. Gayo-Avello, "A survey on session detection methods in query logs and a proposal for future evaluation," Inf. Sci., vol. 179, no. 12, pp. 1822-1843, 2009. https://doi.org/10.1016/j.ins.2009.01.026

[10] D. Downey, S. Dumais, and E. Horvitz, "Models of searching and browsing: languages, studies, and applications," in Proceedings of the 20th international joint conference on Artifical intelligence, Hyderabad, India, 2007, pp. 2740-2747.

[11] F. Radlinski, and T. Joachims, "Query chains: learning to rank from implicit feedback," in Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, Chicago, Illinois, USA, 2005, pp. 239-248. https://doi.org/10.1145/1081870.1081899

[12] M. Munk, and M. Drlik, "Influence of Different Session Timeouts Thresholds on Results of Sequence Rule Analysis in Educational Data Mining," Digital Information and Communication Technology and Its Applications, Communications in Computer and Information Science H. Cherifi, J. Zain and E. El-Qawasmeh, eds., pp. 60-74: Springer Berlin Heidelberg, 2011. https://doi.org/10.1007/978-3-642-21984-9_6

[13] L. D. Catledge, and J. E. Pitkow, "Characterizing browsing strategies in the World-Wide Web," Comput. Netw. ISDN Syst., vol. 27, no. 6, pp. 1065-1073, 1995. https://doi.org/10.1016/0169-7552(95)00043-7

[14] S. Chakrabarti, "Mining the Web's Link Structure," vol. 32, no. 8, pp. 60-67, 1999. https://doi.org/10.1109/2.781636

[15] T. Huynh, and J. Miller, "Empirical observations on the session timeout threshold," Inf. Process. Manage vol. 45, no. 5, pp. 513-528, 2009. https://doi.org/10.1016/j.ipm.2009.04.007

[16] J. Zhang, and A. A. Ghorbani, "The reconstruction of user sessions from a server log using improved time-oriented heuristics." pp. 315-322.

[17] R. Agrawal, T. Imieliski, and A. Swami, "Mining association rules between sets of items in large databases," in Proceedings of the 1993 ACM SIGMOD international conference on Management of data, Washington, D.C., USA, 1993, pp. 207-216. https://doi.org/10.1145/170036.170072

[18] F. Azuaje, "Witten IH, Frank E: Data Mining: Practical Machine Learning Tools and Techniques," BioMedical Engineering OnLine, vol. 5, no. 1, pp. 1-2, 2006/09/29, 2006.

[19] J. Han, L. V. S. Lakshmanan, and J. Pei, "Scalable frequent-pattern mining methods: an overview," in Tutorial notes of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, California, 2001, pp. 5.1-5.61. https://doi.org/10.1145/502786.502792

[20] M. J. Berry, and G. S. Linoff, Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management: John Willey & Sons, 2004.

9 Authors

Jozef Kapusta is with Department of Informatics Constantine the Philosopher University in Nitra, Nitra, Slovakia. He also works at 2Institute of Computer Science, Pedagogical University of Cracow, Cracow, Poland.

Michal Munk is with Department of Informatics Constantine the Philosopher University in Nitra, Nitra, Slovakia. mmunk@ukf.sk

Dominik Halvoník is with Department of Informatics Constantine the Philosopher University in Nitra, Nitra, Slovakia.

Martin Drlík is with Department of Informatics Constantine the Philosopher University in Nitra, Nitra, Slovakia. mdrlik@ukf.sk.