

Review Article

Recommender System for Multiple Databases Based on Web Log Mining

Wan Hussain Wan Ishak* and Nurul Farhana Ismail

Universiti Utara Malaysia, Sintok, Kedah, Malaysia

hussain@uum.edu.my; nurulfana.ismail@gmail.com

*Correspondence: hussain@uum.edu.my

Received: 17th October 2020; Accepted: 15th November 2020; Published: 20th March 2021

Abstract: Finding information from a large collection of resources is a tedious and time-consuming process. Due to information overload, searchers often need help and assistance to search and find the information. Recommender system is one of the innovative solutions to the problem related to information searching and retrieval. It helps and assist searchers by recommending the possible solution based on the previous search activities. These activities can be obtained from the web log, which requires a web log mining approach to extract all the keywords. In this study, keywords obtained from the library web log were analysed and the search keyword patterns were obtained. These keyword patterns were from several databases or resources that were subscribed by the library. The finding revealed some of the popular keywords and the most searchable databases among the searchers. This information was used to design and develop the recommender system that can be used to assist other searchers. The usability test of the recommender system showed that it is beneficial and useful to the searchers. These findings will also benefit the management in planning and managing the subscription of online databases at the university's library.

Keywords: *Information retrieval; recommender system; web log mining; multiple databases; information overload*

1. Introduction

Information is a piece of knowledge that is stored in various forms such as printed and digital forms. Digital forms of storing have been widely implemented to reduce the dependency on printed materials especially paper and ink and to minimise the physical storing space. To date, digital technology has had an impact on economic and environmental performance [1]. Digital materials have many advantages, such as cost reduction, environmentally friendly, easy to disseminate and share through any communication medium, and easy to print when necessary.

Information storing and retrieval in digital form is known as information retrieval (IR) system [2]. The IR system consists of a database system and search interface [3]. The search interface is equipped with a search engine that can be used to retrieve information using queries [4,5]. The information enters the database either through manual or automatic procedure. The size of the database is typically large as the information increases over the years. To date, due to the big data evolution, the traditional IR model needs to be improved [6].

Recommender system (RS) has been used to improve the existing IR systems. RS can help search by recommending items related to user interests [7,8]. RS can assist searchers in finding information and creating the personalized content [9]. For instance, based on the search history, the RS can provide suggestions to the searcher [7,10].

Data mining, specifically web log mining, is one of the approaches that can be used to analyse the web logs by considering the history of the search activity. The log contains data with regard to the search history, database access, and downloads. The log also stores some of the searcher's

information such as Internet Protocol (IPs), sessions, date and time. These data can be utilised by the RS to help other searchers find similar information.

This paper presents the RS model for a multiple independent database system using web log mining. A multiple independent database system is a system that contains a collection of online databases owned or operated by different service providers or publishers. Examples of online databases are ACM Digital Library, EBSCOhost, Emerald, ERIC, JSTOR, ProQuest, etc. Searching in online databases is a very tedious and time-consuming process [11]. Each database may need to be logged in individually and has its own search system. Users will have to switch searching between the search systems. Additionally, users may not be able to determine which database is the most suitable or contain what they want. Therefore, the RS can be implemented as an interface to the multiple independent database system and provide recommendation assistance such as which database should be used.

2. Literature Review

The RS is a user support system for searching and finding information, services, items, or products on the World Wide Web (WWW). It is an application system and method that provides recommendations to users to make decisions on selecting the best items [12]. It has been used in conjunction to the IR system to provide relevant suggestions to users for their collection of items or products that they might be interested in [7,13]. According to Burke [14], the design of the recommendation engine is based on domain and specific features of the available data.

Typically, RS is categorized into three types; collaborative, content-based, and hybrid systems [8,15]. Collaborative filtering generates recommendations by analysing historical interaction. It can be divided into neighbourhood and model-based approaches. The neighbourhood-based approach relies on the rating made by the other users that were selected based on their similarity. The model-based approach employed statistical model to generate the recommendation. The pure collaborative model relies heavily on the user rating matrix. This approach considers all users and items as atomic units. This limitation has been overcome by considering user profile, also known as content-based filtering. This approach evaluates the representation that describe the items and make a comparison with the user interest. Content-based and collaborative recommenders can be combined to leverage the strengths of both approaches. This technique is called a hybrid RS.

Generally, the RS has been implemented in various fields such as entertainment, e-commerce, and services. In e-commerce RS sites were implemented to recommend products to their customers by utilizing their preferences and other customers' purchase history [16]. According to Herlocker et. al [17] there are many advantages of RS in e-commerce. Among the benefits are it helps buyer to find the products that best matches with their needs and reduce the searching and browsing effort.

According to Ricci *et. al* [12], the RS commenced development of a simple observation, in which a person's decision usually depend on the suggestions given by others. Based on their study, they highlighted the importance of the RS as to increase the sold items, sell various items, increase consumer satisfaction, improve consumer well-being and more understanding on the consumers need.

In the tourism field, the RS can be used to assist travellers in finding the destination that best which suits their interests [18]. Alrasheed *et. al* [18] proposed RS which provides users with a bunch of destinations preferred by similar travelers to enable building a list of recommended destinations. Then, the RS ranks the list based on user preferences and constraints. Esmaeili *et. al* [19] proposed a social-hybrid RS for suggesting tourist attractions. Their study utilised several factors such as similarity of traveller's interests, trust, reputation, relationships, and social communities. These factors can increase the RS recommendation quality.

Data mining has good potential in RS research. Data mining is among the steps in knowledge discovery in databases (KDD) [20]. KDD is an automated exploration analysis and modelling of big data repositories [21]. KDD involves a nontrivial process of extracting useful patterns in data [22]. In KDD, data mining is a process consisting of applying data analysis and appropriate algorithms to generate meaningful patterns over a large dataset. For instance, Mohsin *et. al* [23] demonstrates that

the mining of the historical data provides useful patterns that can be used in the present decision making and planning. Furthermore, data mining has been proven to be able to generate good quality of knowledge for decision maker [24].

Etzioni [25] introduced a concept of web mining to find, identify, uncover, and explore information, documents, and knowledge from the web. Etzioni also emphasises that the web mining can support searcher to browse, search, and visualise the web contents. Web mining contains two categories: web content mining and web usage mining [20]. Web content mining is the process of obtaining interesting information from web pages [26]. Web content information can be classified into four types of data information, namely unstructured, structured, semi-structured, and multimedia content [27]. The web usage mining can be applied to find the web usage patterns [26]. It can be used to predict searcher's preferences and behaviour [28].

3. Methodology

This study involved three main phases, which are identifying search keywords and database, frequent keyword analysis, and recommendation modelling.

The first phase comprised three main steps: data selection, data pre-processing, and search keyword detection. Data selection concerned with the selection of appropriate data for this study, while data pre-processing defined a series of actions to be undertaken before the data could be used in the experiment. The search keyword detection aimed to identify the keywords used in the query and its associated database.

In this study, the case study was on library of Universiti Utara Malaysia (UUM). Therefore, web log of the online databases that the library subscribed was obtained. These data were typically secured under the eResources application, which was one of the online services managed by the library. The dataset of server log data in the eResources application under the PSB website used in this study consisted of 32 log files. Each log file contained 500,000 lines.

The data were cleaned and the keywords used by the searchers were identified. The keywords were found in the Uniform Resource Locator (URL) which ended with "query", "keyword", and etc. Figure 1 shows example of the keywords identified in the web log. The keywords would be extracted, and the database associated with that keywords would be stored. As shown in Figure 2, "sukuk+structure", "continuous+education+ and+retention" were among the keywords that were identified from the log. In the next step, all stop words were detected and removed.

```

GET http://www.emeraldinsight.com/80/search.htm?st=sukuk+structure&ct=all&ec=1&bf=1 HTTP/1.1
GET http://proquest.umi.com/80/pgdweb?RQT=302&COPT=REJTPUcyODcnM2mMCLTlQ9MCMZWRVl9lMg==&clientid=28929&cf=1 HTTP/1.1
GET http://www.emeraldinsight.com/80/journals.htm?issn=1753-8394&volume=2&issue=4&articleid=1828941&show=pdf HTTP/1.1
GET http://www.emeraldinsight.com/80/journals.htm?issn=1759-0817&volume=1&issue=1&articleid=1858421&show=pdf HTTP/1.1
GET http://proquest.umi.com/80/pgdweb?SQ=continuous+education+and+retention&DBId=G647&date=ALL&onDate=&beforeDate=&afterDate=&factors=&DBId=G647 HTTP/1.1
GET http://www.emeraldinsight.com/80/search.htm?st=sukuk+structure&ec=1&bf=1&ct=jnl HTTP/1.1
GET http://proquest.umi.com/80/pgdweb?SQ=continuous+education+factors&DBId=G647&date=ALL&onDate=&beforeDate=&afterDate=&factors=&DBId=G647 HTTP/1.1
GET http://eserv.uum.edu.my/80/connect?session=s5Kk6uj4JmXDFogY&url=menu HTTP/1.1
GET http://eserv.uum.edu.my/80/menu HTTP/1.1
GET http://www.emeraldinsight.com/80/journals.htm?issn=0828-8666&volume=26&issue=1&articleid=1846113&show=pdf HTTP/1.1
GET http://www.emeraldinsight.com/80/search.htm?st=sukuk+structure&ec=1&bf=1&ct=jnl&nolog=167626&page=2 HTTP/1.1
GET http://eserv.uum.edu.my/80/public/rightmenu.htm HTTP/1.1
GET http://www.emeraldinsight.com/80/search.htm?st=sukuk+structure&ec=1&bf=1&ct=jnl&nolog=167626&page=2 HTTP/1.1
GET http://www.emeraldinsight.com/80/journals.htm?issn=1753-8394&volume=2&issue=2&articleid=1795358&show=pdf HTTP/1.1
GET http://www.emeraldinsight.com/80/journals.htm?issn=1759-0817&volume=1&issue=1&articleid=1858422&show=pdf HTTP/1.1
GET http://www.emeraldinsight.com/80/search.htm?st=sukuk+structure&ec=1&bf=1&ct=jnl&nolog=167626&page=3 HTTP/1.1
GET http://www.emeraldinsight.com/80/search.htm?st=attitude&ec=1&bf=1&ct=jnl&nolog=814577&page=2 HTTP/1.1
GET http://www.emeraldinsight.com/80/search.htm?st=changing+attitude&ct=all&ec=1&bf=1 HTTP/1.1
GET http://eserv.uum.edu.my/80/connect?session=sjEm3Xln7zCwY3&url=menu HTTP/1.1

```

Figure 1. Sample keywords in a log file

In the second phase, the search keywords that had been obtained from Phase 1 were analysed. The process involved sorting and grouping of the keywords and its associated database. The analysis of frequency was performed using Microsoft Excel to obtain the frequent search keywords.

Finally, the RS model was designed and developed to illustrate the workability of the actual RS. The model was implemented using Hypertext Preprocessor (PHP) language and MySQL database. Figure 2 shows the proposed design of the RS interface. The system acted as an interface to the existing eResources system hosted by PSB UUM. The system was able to suggest which databases contained the keywords that the user had entered.

Recommender System for E-Resources

Enter a keyword:

Most user find articles related to "management" from : Serial Solution

Alternatively, user may find related articles from the following database:

1. ieee
2. ScienceDirect
3. Emerald
4. OnlineLibrary
5. JSTOR
6. SCOPUS
7. Eric
8. Ebscohost
9. JournalCambridge
10. XREF

Figure 2. Proposed design of eResources RS

4. Findings

Based on the analysis, 19,146 keywords were identified and retrieved. In addition, 11 databases that are associated with the keywords were also identified. This finding is depicted in Table 1. Table 1 shows that IEEE database is the most popular where 4,862 keywords were obtained. Of these, 3,418 keywords are unique, while the rest are repetitive. Serial Solutions is the second popular database where 3,043 keywords were obtained. Other databases records less than 2,000 keywords.

Table 1. Database and Number of Keywords

| No. | DATABASE | NUMBER OF KEYWORDS | FREQUENT KEYWORDS | SUM OF KEYWORDS |
|-----|------------------|--------------------|-------------------|-----------------|
| 1. | ebscohost | 1552 | 217 | 1835 |
| 2. | Emerald | 912 | 74 | 1011 |
| 3. | Eric | 462 | 341 | 1081 |
| 4. | ieee | 3418 | 808 | 4862 |
| 5. | JournalCambridge | 1080 | 67 | 1148 |
| 6. | JSTOR | 1321 | 150 | 1503 |
| 7. | OnlineLibrary | 956 | 80 | 1044 |
| 8. | ScienceDirect | 1561 | 255 | 1906 |
| 9. | SCOPUS | 1226 | 124 | 1379 |
| 10. | SerialSolution | 2408 | 363 | 3043 |
| 11. | XREF | 307 | 18 | 334 |
| | GRAND TOTAL | 6143 | 2497 | 19146 |

The analysis of the keywords frequency together with its associated databases is depicted in Table 2. The table illustrated that most of the keywords were used twice and thrice. Databases, such as IEEE, Serial Solutions, ScienceDirect, EBSCOhost, and ERIC, contained keywords that repeated more than three times. The IEEE database had nine keywords that repeated more than ten times, while Serial Solutions had five keywords that repeated more than 10 times.

Table 3 demonstrates the example analysis of the 30 highest frequent keywords that were identified from the databases. Keywords such as "management", "performance", "entrepreneur", "system", "service", "business", "theory", and "relationship" were used in all databases. Keyword "management" was the most popular where it was searched 68 times in all databases. While "methodology" is less sought by the searchers.

The usability test was also conducted on the recommender search interface. Based on the feedback, most of the respondents stated that they were familiar with the functionalities of online database system. About 90% of users agreed that they needed guidance when using the prototype for the first time, while the other 10% said the system was easy to use and user friendly. It showed

that all the users agreed that the existing eResource module did not have all the necessary functions and capabilities as they expected. Most of the users expected the eResource module to have the capabilities in helping users on searching the information. They also agreed on having a RS for this eResource module. Overall, most of the users stated that the proposed RS interface was simple and easy to use. They also agree that the interface and information provided are clear and effective in supporting the searchers.

Table 2. Databases and Keyword Frequency

| NO. | DATABASE | KEYWORD FREQUENCIES | | | | | | | | | |
|-------|------------------|---------------------|------|-----|-----|----|----|----|---|---|-----|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | >10 |
| 1 | ebscohost | 1335 | 168 | 40 | 6 | 1 | 1 | 0 | 0 | 1 | 0 |
| 2 | Emerald | 838 | 54 | 16 | 3 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | Eric | 121 | 121 | 194 | 7 | 8 | 10 | 0 | 1 | 0 | 0 |
| 4 | ieee | 2610 | 516 | 151 | 60 | 39 | 18 | 9 | 3 | 3 | 9 |
| 5 | JournalCambridge | 1013 | 66 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | JSTOR | 1171 | 130 | 11 | 6 | 3 | 0 | 0 | 0 | 0 | 0 |
| 7 | OnlineLibrary | 876 | 75 | 3 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 8 | ScienceDirect | 1306 | 196 | 41 | 11 | 3 | 2 | 2 | 0 | 0 | 0 |
| 9 | SCOPUS | 1102 | 107 | 8 | 6 | 3 | 0 | 0 | 0 | 0 | 0 |
| 10 | SerialSolution | 2045 | 240 | 59 | 36 | 9 | 9 | 3 | 1 | 1 | 5 |
| 11 | XREF | 289 | 12 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| TOTAL | | 12706 | 1685 | 527 | 139 | 68 | 40 | 14 | 5 | 5 | 14 |

Table 3. The 30 most popular keywords

| NO. | KEYWORD | NUMBER OF SOURCE DATABASES | KEYWORD FREQUENCY | | | | |
|-----|------------------|----------------------------|-------------------|-----|-----|---------|--------|
| | | | TOTAL | MIN | MAX | AVERAGE | MEDIAN |
| 1. | management | 11 | 68 | 1 | 20 | 6.18 | 4 |
| 2. | organization | 9 | 44 | 2 | 12 | 4.89 | 5 |
| 3. | network | 10 | 39 | 1 | 15 | 3.90 | 2 |
| 4. | technology | 10 | 37 | 1 | 10 | 3.70 | 2 |
| 5. | performance | 11 | 36 | 1 | 7 | 3.27 | 3 |
| 6. | entrepreneur | 11 | 35 | 1 | 10 | 3.18 | 2 |
| 7. | communication | 10 | 34 | 1 | 14 | 3.40 | 2 |
| 8. | organizational | 10 | 34 | 1 | 6 | 3.40 | 3 |
| 9. | information | 10 | 32 | 1 | 9 | 3.20 | 2.5 |
| 10. | system | 11 | 32 | 1 | 11 | 2.91 | 2 |
| 11. | service | 11 | 31 | 1 | 7 | 2.82 | 3 |
| 12. | business | 11 | 30 | 1 | 8 | 2.73 | 2 |
| 13. | company | 10 | 30 | 1 | 10 | 3.00 | 2 |
| 14. | application | 10 | 29 | 1 | 10 | 2.90 | 2.5 |
| 15. | entrepreneurship | 10 | 29 | 1 | 9 | 2.90 | 1.5 |
| 16. | attitude | 10 | 28 | 2 | 4 | 2.80 | 2.5 |
| 17. | characteristic | 9 | 28 | 1 | 4 | 3.11 | 3 |
| 18. | strategy | 10 | 28 | 1 | 6 | 2.80 | 2 |
| 19. | theory | 11 | 28 | 1 | 6 | 2.55 | 2 |
| 20. | journal | 10 | 27 | 1 | 6 | 2.70 | 2.5 |
| 21. | Knowledge | 9 | 27 | 1 | 6 | 3.00 | 2 |
| 22. | Malaysia | 9 | 27 | 1 | 10 | 3.00 | 2 |
| 23. | student | 10 | 26 | 1 | 6 | 2.60 | 2 |
| 24. | difference | 10 | 25 | 1 | 5 | 2.50 | 2 |
| 25. | effect | 10 | 25 | 1 | 6 | 2.50 | 2 |
| 26. | employee | 10 | 25 | 1 | 5 | 2.50 | 2 |
| 27. | impact | 10 | 25 | 1 | 5 | 2.50 | 2 |
| 28. | relationship | 11 | 25 | 1 | 4 | 2.27 | 2 |
| 29. | development | 10 | 24 | 1 | 6 | 2.40 | 2 |
| 30. | methodology | 7 | 24 | 1 | 11 | 3.43 | 2 |

5. Conclusion

This study proves that library users actively use online databases to find and obtain information for academic research purposes. The patterns reveal that users are searching for information from more than one databases. This can be seen by the use of 30 most popular keywords in most of the databases. The proposed RS interface was found to be beneficial to the users, assisting them in finding the information in the "right" databases.

The major constraint faced by this study is the data size. The data size was too big, too difficult, and took a great deal of effort to process. Furthermore, the raw data contained too much “rubbish”. The quality of the data was affected and some of the data might be lost during the processing.

Based on the nature and size of the web log data, an automatic mechanism for extracting useful data needs to be designed and developed. The mechanism should be able to distinguish the major fields such as keywords, database name, session, and user profile. Therefore, a specific study should embark on this automatic extraction. It will be a useful tool for research in this field.

Future studies should also include user profile as one of the information to be extracted. User profile can increase the recommendation as those who have a similar profile might have the same interest.

Acknowledgements

This work was financially supported by Universiti Utara Malaysia (UUM) under the university grant funding. Additional thanks to the Research and Innovation Management Centre (RIMC) for the technical support and Perpustakaan Sultanah Bahiyah, the library of UUM, for sharing the web log data.

References

- [1] Li, Y., Dai, J. and Cui, L. (2020). The Impact of Digital Technologies on Economic and Environmental Performance in the Context of Industry 4.0: A Moderated Mediation Model. *International Journal of Production Economics*, 229, 107777, <https://doi.org/10.1016/j.ijpe.2020.107777>.
- [2] Marchionini, G. and White, R. (2007). Find What You need, Understand What You Find. *International Journal of Human-Computer Interaction*, 23(3), 205-237.
- [3] Yamin, F.M. and Ramayah, T. (2013). Interfacing Google search engine to capture user web search behavior, *International Journal of Electronic Commerce Studies*, 4(1), 47-62.
- [4] Day, J. (2001). The Quest for Information: A Guide to Searching in the Internet. *Journal of Contemporary Dental Practice*, Nov: (2)4, 33-43.
- [5] Yamin, F.M., Ramayah, T. and Ishak, W.H.W. (2013). Information Searching: The Impact of User Knowledge on User Search Behavior. *J. Info. Know. Mgmt.*, 12(3), 1350023, <https://doi.org/10.1142/S0219649213500238>
- [6] Wang, C., Qin, F. and Samuel, D.K. (2020). Cloud Assisted Big Data Information Retrieval System for Critical Data Supervision in Disaster Regions. *Computer Communications*, 151, 548-555, <https://doi.org/10.1016/j.comcom.2019.11.028>.
- [7] Walek, B. and Fojtik, V. (2020). A hybrid Recommender System for Recommending Relevant Movies using an Expert System. *Expert Systems with Applications*, 158, 113452, <https://doi.org/10.1016/j.eswa.2020.113452>.
- [8] Melville, P. and Sindhvani, V. (2010). Recommender Systems. In C. Sammut, G. Webb (eds.), *Encyclopedia of Machine Learning*, Springer-Verlag Berlin Heidelberg, pp: 1-9.
- [9] Mahmood, T., Ricci, F. and Venturini, A. (2009). Learning adaptive recommendation strategies for online travel planning. *Inf. Commun. Technol. Tour.* 2009, pp. 149–160
- [10] Grossman, L. (2010). How Computers Know What We Want — Before We Do, *Time* (Thursday, May 27, 2010). Available: <http://www.time.com/time/printout/0,8816,1992403,00.html>
- [11] Roy, P., Kumar, S. and Satija, M.P. (2012). Problems in Searching Online Databases: A Case Study of Select Central University Libraries in India. *DESIDOC Journal of Library & Information Technology*, 32(1), 59-63.
- [12] Ricci, F., Rokach, L. and Shapira, B. (2015). *Recommender Systems Handbook* (2nd). New York: Springer
- [13] Sindhvani, P.M.V. (2010) Recommender systems. *Commun. ACM*, pp. 1–21
- [14] Burke, R. (2007). Hybrid Web Recommender Systems. *The Adaptive Web*, pp. 377–408
- [15] Adomavicius, G. and Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transaction on Knowledge and Data Engineering*, 17(6), 734-749.
- [16] Montaner, M., López, B. and Rosa, J.L. (2003) A Taxonomy of Recommender Agents on the Internet. *Artificial Intelligence Review*, 19, 285–330.
- [17] Herlocker, J. L., Konstan, J. A., Terveen, L. G. and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1), pp. 5–53.
- [18] Alrasheed, H., Alzeer, A., Alhowimel, A., Shameri, N. and Althyabi, A. (2020). A Multi-Level Tourism Destination Recommender System. *Procedia Computer Science*, 170, 333-340, <https://doi.org/10.1016/j.procs.2020.03.047>.

- [19] Esmaeili, L., Mardani, S., Golpayegani, S.A.H. and Madar, Z.Z. (2020.) A Novel Tourism Recommender System in the Context of Social Commerce. *Expert Systems with Applications*, 149, 113301, <https://doi.org/10.1016/j.eswa.2020.113301>.
- [20] Cooley, R., Mobasher, B. and Srivastava, J. (1997). Web mining: information and pattern discovery on the World Wide Web. *Proceedings Ninth IEEE International Conference on Tools with Artificial Intelligence*, Newport Beach, CA, USA, pp. 558-567, DOI: 10.1109/TAI.1997.632303.
- [21] Maimon, O. and Rokach, L. (2005) Introduction to knowledge discovery in database. *Data Min. Knowl. Discov. Handb.*, pp. 1–17.
- [22] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Mag.*, pp. 37–54.
- [23] Mohsin, M. F. M., Hibaullah, C. F., Norwawi, N. M. and Wahab, M. H. A. (2010). Mining the student programming performance using rough set, *IEEE International Conference on Intelligent Systems and Knowledge Engineering*, Hangzhou, pp. 478-483, doi: 10.1109/ISKE.2010.5680824.
- [24] Mohsin, M. F. M. and Wahab, M. H. A. (2008). Comparing the knowledge quality in rough classifier and decision tree classifier, *International Symposium on Information Technology*, Kuala Lumpur, 2008, pp. 1-6, doi: 10.1109/ITSIM.2008.4631700.
- [25] Etzioni, O. (1996). The World-Wide Web: Quagmire or Gold Mine?, *Commun. ACM*, 39(11), pp. 65–68.
- [26] Srivastava, J., Cooley, R., Deshpande, M. and Tan, P-N. (2000). Web usage mining: discovery and applications of usage patterns from Web data. *SIGKDD Explor. Newsl.* 1, 2 (January 2000), 12–23., DOI: 10.1145/846183.846188.
- [27] Adsod, A. A. and Chopde, N. R. (2014). A Review on: Web Mining Techniques. *Int. J. Eng. Trends Technol.*, 10(3), pp. 108–113.
- [28] Mobasher, B., Cooley, R. and Srivastava, J. (2000). Automatic Personalization Based on Web Usage Mining. *Communications of the ACM*, 43(8), pp: 142-151.



© 2021 by the author(s). Published by Annals of Emerging Technologies in Computing (AETiC), under the terms and conditions of the Creative Commons Attribution (CC BY) license which can be accessed at <http://creativecommons.org/licenses/by/4.0>.