# Models of virtual library users' behavior analysis

## Gytis Vievesis*[a]* , Vytautas Janilionis*[b]* , Antanas Štreimikis*[a]*

[a] *Department of Information Technology Services, Kaunas University of Technology*
 Studentų 48A, LT-51367, Kaunas

[b] *Department of Applied Mathematics, Kaunas University of Technology*
 Studentų 50, LT-51368, Kaunas

 E-mail: gytis.vievesis@ktu.lt; vytautas.janilionis@ktu.lt; antanas.streimikis@ktu.lt

**Abstract.** In this paper, we present models for the analysis of the behavior of the virtual library (VL) users. Unlike the models presented in the literature, they use only the big data that is stored in the log files of virtual library servers and methods of statistics, association rules, and recommendation systems. The proposed models were implemented with R software. Using the proposed models, the analysis of the behavior of VL users of Lithuanian research and study of higher education institutions was performed for the first time. The results showed that the proposed models allow to operatively analyze the behavior of virtual library users using advanced search filters, facets, and provide suggestions for improvement of service quality.

**Keywords:** virtual library; behavior of users; statistics; big data; association rules; recommender systems

**AMS Subject Classification:** 62P30

## 1 Introduction

When different types of virtual library services are provided online, it is important to ensure high-quality services for different VL users. A review of the scientific literature has revealed that the configuration of the layout of facets and filters is usually based on surveys or intuition of the librarians and administrators. Such a layout of VL search filters and facets does not fully meet the real needs of VL users and the growing needs for service quality are deficiently satisfied. The additional solutions need to be integrated for more efficient use of search and facet into VL to find the necessary

sources as soon as possible. The analysis of search filters and facets could allow us to make recommendations on how to significantly speed up the filtering of search results and improve the quality of the VL services.

Analysis of user behavior is performed by applying the methods of classification, clustering, visualization [7]. Most often the analysis of users' behavior is executed via interviews and surveys. According to findings, web logs analysis allows not only to identify actions of VL users but also to accumulate detailed information, e.g. about the search words, advanced search filters, and facets (filters applied for search) [1]. The literature [8] highlights the three most popular results filters – Resource Type, Creation Date, Topic. According to findings, the results filters placed on the top of the VL webpage have been used most frequently. Libraries are increasingly focusing on web analytics when data is collected automatically and reflects the actual website users' actions [4]. This is the way to identify users' actions, yet the reasons for such behavior remain unclear. It is important to take account of the needs of different VL users and their most frequently used facets and filters. It is also necessary to ensure that users could choose facets that are relevant for them and the place in the VL webpage is appropriate.

## 2 The models for the analysis of VL users' behavior

The proposed models of users' behavior research consist of the exploratory analysis of users' behavior, analysis of filters and facets, and building recommender systems. Analysis of filters and facets allows determining which facets and filters are used most frequently or simultaneously. To perform the advanced search filters and facets analysis, it is not necessary to conduct surveys, interviews, or additional software, only log files is proposed to use. Parsing of URL by the parameters, as well as the development of the lists of used filters and facets for each action of search is necessary. This data allows generating the association rules defining the relations between two sets of search filters and facets.

Apriori method [2] is used to identify association rules. It follows an iterative approach commonly known as a level wise-search, where $k$-itemsets are used to explore $(k + 1)$-itemsets. At first, the set of frequent 1-itemsets, denoted as $L_1$, is determined. $L_1$ is then used to find the set of frequent 2-itemsets $L_2$ and so on until no more frequent $k$-itemsets can be determined. Let $I = \{i_1, i_2, \ldots, i_m\}$ denote as the set of $m$ facets. A rule is defined as an implication of the form $X \Rightarrow Y$ where $X, Y \subseteq I$ [2]. Set of facets $X$ and $Y$ are called the left-hand side (LHS) and right-hand side (RHS) of the rule [2]. Support

$$s(X \Rightarrow Y) = P(X \cap Y) \cdot 100\%, \tag{1}$$

confidence

$$c(X \Rightarrow Y) = \frac{P(X \cap Y) \cdot 100\%}{P(X)}, \tag{2}$$

and lift

$$l(X \Rightarrow Y) = \frac{P(X \cap Y)}{P(X)P(Y)} \tag{3}$$

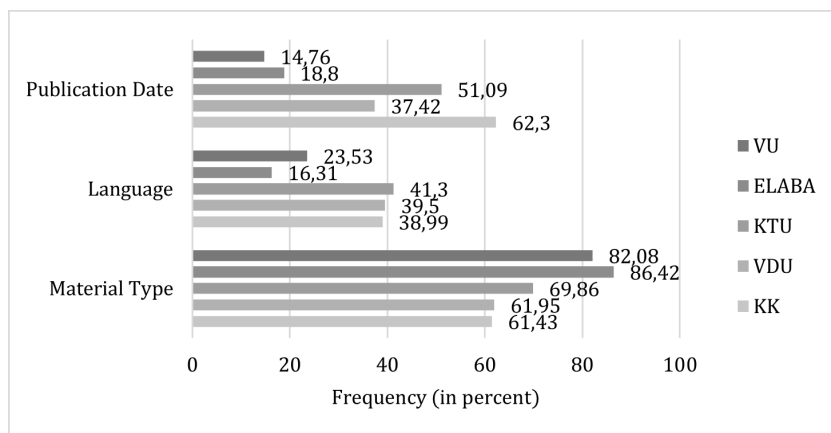are used to evaluate the association rules [2, 6].

**Fig. 1.** Frequencies of the advanced search filters usage.

A recommender system is suggested to recommend additional facets for VL users. Three methods are used to implement recommender systems: Random Items, Popular Items, and Item-Based Collaborative Filtering (IBCF). IBCF method recommends items that are similar to the items that users prefer [5]. The prediction of the user $i$ to the item $j$ is

$$p_{i,j} = \lambda \sum_{k=1}^{m} w(k,j) v_{i,k}, \tag{4}$$

where $w(k,j)$-similarity of items calculated by cosine similarity of users; $v_{i,k}$ – user $i$ rating to item $k$; $\lambda$ is the normalization factor [3].

Receiver Operating Characteristic curve (ROC) is used to identify the most accurate method [9]. ROC curve is a plot of the recommender system's probability of detection (true positive rate (TPR)) by the probability of false alarm (false positive rate (FPR)) [5]. R programming language was used for the implementation of the models.

## 3 Analysis of VL users' behavior of Lithuanian higher education institutions

Using the proposed models, the analysis of the virtual library (based on Ex Libris Primo search and discovery tools) users' behavior of Lithuanian research and higher education institutions was performed for the first time. The big data of server logs of May 2020 were analyzed and accumulated 12.3 million log events during this period. 828.000 log events were selected complying with the users' actions. During the period under analysis 51,000 unique VL users have been identified.

The analysis of advanced search filters and facets was performed to analyze the frequency of their usage, which filters, and facets are used at the same time, and how frequently their combinations occur in the searches of VL users. Three advanced search filters are used in VL: Material Type, Publication Date, Language. Figure 1 presents the frequencies of filters' usage in five VL: Vilnius University (VU),

**Table 1.** Association rules of advanced search filters.

| Rule | | | Support | Confidence | Lift |
|---|---|---|---|---|---|
| LHS | | RHS | | | |
| {Language} | → | {Publication Date} | 10.5% | 48.3% | 2.0 |
| {Publication Date} | → | {Language} | 10.5% | 43.7% | 2.0 |
| {Material Type, Language} | → | {Publication Date} | 7.4% | 51.1% | 2.1 |
| {Material Type, Publication Date} | → | {Language} | 7.4% | 55.3% | 2.5 |

**Table 2.** Association rules of facets.

| Rule | | | Support | Confidence | Lift |
|---|---|---|---|---|---|
| LHS | | RHS | | | |
| {eLABa Institution, Language} | → | {Access Rights of eLABa Object} | 0.1% | 18.3% | 12.6 |
| {Access Rights of eLABa Object, Language} | → | {eLABa Institution} | 0.1% | 35.8% | 7.8 |
| {Resource Type, Availability, Topic} | → | {Language} | 0.3% | 67.0% | 7.0 |
| {Resource Type, Access Rights of eLABa Object} | → | {eLABa Institution} | 0.2% | 29.7% | 6.5 |

Lithuanian Academic Electronic Library (ELABA), Kaunas University of Technology (KTU), Vytautas Magnus University (VDU), Kaunas University of Applied Sciences (KK). Only the advanced searches with filters were analyzed. According to the findings, the most popular filter is Material Type.

The association rules for usage of advanced search filters have been created and their analysis carried out. Table 1 presents the association rules, support, confidence, and lift of them. Association rules of advanced search filters were generated and four of them have lift value greater than 1. For example, the first rule shows that both filters are used in 10.5 percent of advanced searches. The lift value greater than 1 (lift = 2) shows that the occurrence of filter Language has a positive effect on the occurrence of the publication date in the advanced searches with filters.

About 30 facets are used in VL. The most frequently used facets are Resource Type, Availability, Creation Date, Language, eLABa institution. Other facets are used infrequently, e.g.: eResource Collection, FMT, eLABa Object Type, New Rocords. They occur in less than 0.7 percent of searches with facets.

Over 200 association rules were generated with the parameters of minimum confidence 0.1 and minimum support 0.1 and the rules with the highest lift are presented in Table 2. For example, filters eLABa Institution, Language, and Access Rights of eLABa Object were used in 0.1 percent of searches with facets, and the occurrence of both filters eLABa Institution and Language has a positive effect on the occurrence of the filter Access Rights of eLABa Object.

To update VL facets, the recommender system for facets was developed using the data of searches when a user applied at least one facet. Three methods were used for the development of the recommender system: Random items, Popular items, and IBCF. Figure 2 presents ROC curve with the TRP and FPR values by applying three different methods for the development of the recommender system.
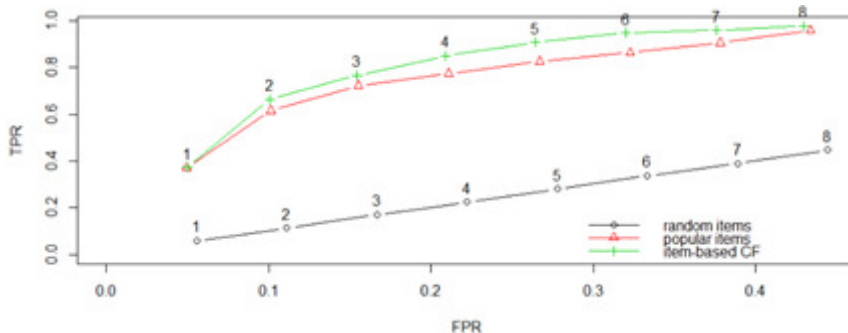
**Fig. 2.** Comparison of three recommender methods.

**Table 3.** Recommended filters.

| Used facets: | Resource Type | eLABa Institution Topic | | Availability |
|---|---|---|---|---|
| Additional facets: | Creation Date Availability Language | Resource Type Journal Title Access Rights of eLABa Object | | Resource Type Creation Date Language |

The obtained results showed that IBCF method for the development of recommender system is the most accurate. Table 3 presents three examples of recommendations based on users' behavior on May 2020, i.e. which facets were used, and additional filters were recommended. For example, if the user has chosen Resource Type results filter, it is recommended to use three additional filters: Creation Date, Availability and Language.

## 4   Conclusions

Analysis of VL users' behavior is very important to improve quality of services for VL users. In this paper, we have proposed two models for the analysis of users' behavior, applying the data of users' behavior collected only from the automatically accumulated server Logs for the first time in VL. The advantage is that the analysis of VL users' behavior does not require user surveys or additional software recording behavior of users. Advanced search filters and facets analysis showed that the behavior of VL users in Lithuanian higher education institutions differs. Association rules revealed differences and similarities in the behavior of various VL users. According to the results of the Recommender System, it was suggested to further develop the layout of VL facets, taking account of the behavior of different users to match the facets with the needs of separate VL users, including a proper layout of facets in the webpage. Proposed models allow to simplify the search filters and facets analysis, investigate the users' behavioral patterns, adapt, and update the content of VL.

## References

[1] T. Bogaard, L. Hollink, J. Wielemaker, J. van Ossenbruggen, L. Hardman. Metadata categorization for identifying search patterns in a digital library. *J. Document.*, **75**(2), 2019.

[2] G. D'Angelo, S. Rampone, F. Palmieri. Developing a trust model for pervasive computing based on Apriori association rules learning and Bayesian classification. *Soft Comp.*, **21**(21):6297–6315, 2017. https://doi.org/10.1007/s00500-016-2183-1.

[3] Y. Dou, H. Yang, X. Deng. A survey of collaborative filtering algorithms for social recommender systems. In *12th International Conference on Semantics, Knowledge and Grids (SKG)*, pp. 40–46, 2016. https://doi.org/10.1109/SKG.2016.014.

[4] J.C. Fagan. The suitability of web analytics key performance indicators in the academic library environment. *J. Acad. Libr.*, **40**(1):25–34, 2014. https://doi.org/10.1016/j.acalib.2013.06.005.

[5] M. Hahsler. *recommenderlab: A Framework for Developing and Testing Recommendation Algorithms.* 2015.

[6] N. Hussein, A. Alashqur, B. Sowan. Using the interestingness measure lift to generate association rules. *J. Adv. Comp. Sci. Techn.*, **4**(1):156, 2015.

[7] P. Suthar, B. Oza. A survey of web usage mining techniques. *Int. J. Comp. Sci. Inf. Techn.*, **6**(6):5073–5076, 2015.

[8] D. Wells. Library discovery systems and their users: a case study from Curtin university library. *Australian Acad. Res. Libr.*, **47**(2):1–14, 2016. https://doi.org/10.1080/00048623.2016.1187249.

[9] J. Wu, A. Martin, R. Kacker. Bootstrap variability studies in ROC analysis on large datasets. *Comm. Stat.: Simulat. Comp.*, **43**(1):225–236, 2014. https://doi.org/10.1080/03610918.2012.700362.

REZIUMĖ

## Virtualių bibliotekų naudotojų elgsenos analizės modeliai

*G. Vievesis, V. Janilionis, A. Štreimikis*

Šiame darbe pristatomi modeliai skirti virtualiųjų bibliotekų naudotojų elgsenos analizei. Skirtingai nuo literatūroje pateiktų modelių, pasiūlytuose modeliuose naudojami tik didieji duomenys, kurie kaupiami virtualiųjų bibliotekų serverių žurnaluose, o analizei taikomi statistikos, susietumo taisyklių ir rekomendacinių sistemų sudarymo metodai. Pasiūlyti modeliai realizuoti su R programine įranga. Panaudojus pasiūlytus modelius, pirmą kartą atliktas Lietuvos mokslo ir studijų institucijų virtualių bibliotekų naudotojų elgsenos tyrimas, kuris parodė, kad pasiūlyti modeliai leidžia operatyviai vertinti virtualių bibliotekų naudotojų elgsenos ypatumus naudojantis paieškos ir rezultatų filtrais bei teikti siūlymus paslaugų kokybės tobulinimui.

*Raktiniai žodžiai*: virtuali biblioteka; naudotojų elgsena; statistika; didieji duomenys; susietumo taisyklės; rekomendacinės sistemos