

Bibliomining Analysis Method - The Application in Economy of Public Sector

P. Hájek, J. Stejskal

Abstract—This paper aims to design the solution for the measurement of the public services in theory and is a practical application of clustering methods and methods of extraction of attributes from real public library data in order to uncover similarities in the services provided by public libraries, based on a representative set of behavior patterns of users. The results of this work can then be implemented to achieve more efficient management of the public library.

Key-Words—bibliomining method, cluster analysis, library, principal component analysis, public library services.

1 INTRODUCTION

MANAGEMENT and marketing options in the organization of the public sector are significantly different to a normal commercial enterprise. This is mainly due to the wide range of differences in the individual sectors - public and private. Some of the major differences are in the mode of financing (from public funds), lack of standards and performance indicators (public goods and services are difficult to measure both in quantity and in quality), it is not possible to use standard proven management models and practices from the private sector and many others. On the other hand, there is currently a great deal of pressure on effective management and process optimization in public sector organizations. It carries with it increased demands for various kinds of analyses that will provide sufficient information for quality management and for achieving the required level of performance (and efficiency).

To achieve this, it is important to prepare a detailed analysis of the portfolio of performance; in the public sector this most often means public services. Usually this is only an extensive list of items (e.g. a price list). This type of list may be very extensive but it does not usually have any added value. Therefore, it is necessary to perform a more detailed unit processing, or classification of the individual services into groups based on their common attributes.

J. Stejskal is now with the Institute of Economics Sciences, Faculty of Economics and Administration, University of Pardubice, Studentská 84, 532 10 Pardubice, Czech Republic (phone: +420466 036067; e-mail: jan.stejskal@upce.cz).

P. Hájek is now with the Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Studentská 84, 532 10 Pardubice, Czech Republic (phone: +420466 036067; e-mail: petr.hajek@upce.cz).

This allows management a more precise focus on the items and a better targeting of organizational strategy, and setting up of processes for the effective delivery of public services.

Public libraries provide a typical example of an area where it is necessary to perform a detailed analysis of the outputs. It is generally known what services a library provides, their list can be found on their website. For management there are many small items which cannot be processed properly - typically it is not possible to monitor and manage the costs enabling better targeting of marketing or distribution of services in the library building itself. For this reason, libraries perform detailed performance analyses but the central methodology in this field is missing. They are not based on empirical data collected from their customers - readers, making it impossible to verify the results. Therefore, outputs of analysis without empirical verification are not accurate and the desired results do not materialize. Bibliomining can be a suitable method of processing empirical data in the sense of a detailed analysis of the outcomes (public services).

Bibliomining is defined in the framework of the provision of library services as the process of applying data mining techniques to extract patterns of behavior from library databases [1,2]. This process consists of several steps, namely, target identification, collection and preprocessing of data sets, discovery of knowledge in data, result evaluation, and application of the acquired knowledge in practice. The aim of this process is to obtain such behavioral patterns of library users, enabling more efficient customer relationship management and continued improvement in the quality of services offered by libraries. On the other hand, these results can be used for effective marketing libraries and other possible personalization of services, e.g. recommendation of items based on the analysis of similar users [3]. Bibliomining is becoming an important tool for support the decision making of library management.

In the past, research in the area of bibliomining has only focused on defining the concept [2] or the technological background of the application, which concerns the process of designing a suitable database [4]. Bibliomining applications have hitherto been focused on the optimal allocation of budgetary resources based on past circulation of documents [5], web usage mining [6,7], analysis of bibliometric data [8] or text mining [9]. What is lacking is a specific application of bibliomining

methods based on the recognition of patterns of user behavior. This work aims to fill this gap and presents the application of clustering methods and methods of extracting attributes of real data of the public library in order to find similarities in the services provided by public libraries based on a representative set of behavior patterns of users. The results of this work can be implemented to achieve more efficient management of the public library.

The rest of the paper is structured as follows. Section 2 is devoted to the characteristics of library services. There is then a description of data collection and processing. Section 4 characterizes the use of bibliomining methods and the results are interpreted. The paper ends with a summary of the conclusions.

2 TYPOLOGY OF LIBRARY SERVICES

It should be noted at the onset that there is no traditional typology or taxonomy of library services in the world or even in Europe. Individual countries may give typology the form of legislation, however this is not common. The only regulation which deals with typology is the IFLA/UNESCO guideline for development from 2002. This defines a non-exhaustive list of basic library services in public libraries as follows:

- loan of books and other media,
- provision of books and other materials for use in the library,
- information services using print and electronic media,
- readers' advisory services including reservation services,
- community information services,
- user education including support for literacy programmes,
- programmes and events,
- remote access to electronic collections and library services,
- and other services that reflect the needs of the local community.

It is clear that due to its excessive brevity processing performed in this way is insufficient. However, it can serve as comparative material for the purposes of clustering individual services into groups.

There is currently no typology of library services either in the Czech Republic or abroad. An example may be the attempt of Municipal Library of Prague (MLP) (the largest public library in the country) to create one.

Services provided by the MLP can be divided into primary (document services), secondary (bibliographic retrieval and reference services) and others (tertiary) services. Some of the services provided by the library are accompanied by other services, i.e. assistance of a librarian.

Primary services can be further divided into loan service (in-house and charge-out, both with assistance if

required), interlibrary loan services, reprographics (photocopying) and digital terminal services (e-documents, full-text database) or remote delivery of documents.

Secondary services include bibliographic search, reference services, documentographic, factographic services and bibliographic information services. Reference services provide information through other sources (identifying location of documents in larger libraries; cooperative information system "Ask a Librarian").

Other services include advisory services (provision of individual consultation and information, and instruction on library services, catalogues, databases, library funds, how to cite sources used, etc.), civil services (from Opencard through elections, petitions), presentation of the library (library website, organizing excursions to the library, instruction on the library and its services, document retrieval, using catalogues, databases and related digital services, etc.), electronic and technical services (provision of free internet access, wireless internet, use of playback equipment or readers in special departments, etc.), publishing services (own publications with ISBN, creativity support, prints, audial, audiovisual, etc.), promotion of reading, cultural and educational events and visitor services (ensuring a pleasant stay in the library other than time spent studying, provision of social facilities and recreation, such as playing games, social communication, etc.).

There are other so-called special services but the MLP does not provide them. These are provided by libraries with a narrow focus or libraries with specialized departments. These services are provided to researchers for their research activities. The library also provides services to users with various disabilities (e.g. visually impaired) or specialization in one particular field or topic. The final wording of the typology prepared by the MLP is included in Appendix 1. This typology should be verified by an empirical study performed using bibliomining methods.

3 DATA COLLECTION AND PRE-PROCESSING

The survey was carried out in the pilot phase of the project Methodology for measuring the value of library services. It was a purely quantitative questionnaire survey conducted to verify the proposed method of determining empirical data among its respondents. The partial aim was to provide enough empirical data for verification of method of questioning the respondents on their individual preferences of the public services provided by the library, including questions on the valuation of their importance or value.

The basic set of respondents was made up of readers at the MLP. The total number of members of the questionnaire panel was 1 061, a total of 620 responses were processed. Among the panel members were

randomly selected MKP readers who meet the following criteria:

- age 15 + (1994 and older)
- loaned at least one item in 2010,
- gave an email address,
- responded in the first round of questioning, thus agreed to be included in the panel.

The panel members were surveyed using an online questionnaire (CAWI type) during October and November 2011. In the pilot survey the panel of respondents was divided into 7 groups, who provided with different variants of the same questions (different concepts of the questions or different questions focusing on the same area of preference, etc.). The method of questioning comes from foreign studies [10-14], which use WTA and WTP procedures to determine the respondents' opinions on the value of the services (for an explanation see [10]). Part of the questions was designed separately by the authors of this paper.

Input attributes were related to the basic socio-economic and demographic characteristics (gender, age, economic activity and education) of the reader, frequency of visits and loans and finally customer behavior, i.e. whether the individual uses the individual services offered by the library. An overview of the attributes relating to the readers' user behavior is shown in Table 1.

This table shows an overview of library services and the frequency of their use by the reference sample of readers. More than 88% of readers use the service loan of books and other media without the assistance of a librarian. This assistance, however, is utilized by 52% of users. Other services are used by less than half of the users. Approximately 43% of users spend time in the library reading documents retrieved without the assistance of a librarian. Less than 10% of users utilize the service of copying and printing of documents. A similar situation is observed in the case of reading and downloading of electronic documents. Half of the readers use the library's website to search for documents in the catalogue. About 38% need the services of a librarian to search. None of the users utilized the possibility of processing research documents and only 2% used the services "Ask a Librarian". Assistance of a librarian is mainly used for the purpose of obtaining information on the availability of books in another branch or library (32%) and instruction on library services (20%). Approximately 17% of users participated in an educational or cultural program in the library, and 7% outside the library. About 41% of users perceive their stay in the library as a form of relaxation and 12% as a place to meet friends. In order to access the Internet, users utilize the library's computers (16%) in approximately the same extent as they use their own computers through Wi-Fi (17%).

Table 1: Reader attributes

Abbrev.	Characteristic	Frequency of usage (yes/no)
x_1	Loan of books, magazines, CD retrieved without the help of a librarian	547/73
x_2	Loan of books, magazines, CDs from stock, retrieved with the help of a librarian	320/300
x_3	Reading books, newspapers or magazines, etc. studying documents etc. in the library - retrieved without the help of a librarian	269/351
x_4	Reading of books, newspapers or magazines, etc. study of documents etc. in the library – retrieved with the help of a librarian	104/516
x_5	Copying of documents (without the assistance of a librarian)	60/560
x_6	Copying of documents (with the assistance of a librarian)	27/593
x_7	Printing of documents (without the assistance of a librarian)	14/606
x_8	Printing of documents (with the help of a librarian)	22/598
x_9	Use of electronic databases	25/595
x_{10}	Reading e-books	5/615
x_{11}	Downloading e-books from the website	59/561
x_{12}	Downloading other electronic documents from the website	20/600
x_{13}	Search in the paper catalogue	21/599
x_{14}	Search in the electronic catalogue on a computer	285/335
x_{15}	Search in the electronic catalogue on the website	308/312
x_{16}	Search with assistance of a librarian	238/382
x_{17}	Processing of written research	0/620
x_{18}	Information on the availability of a book at another branch or library	201/419
x_{19}	Asking a librarian for instruction on library services	124/496
x_{20}	Use of the service "Ask a librarian"	12/608
x_{21}	Participation in an educational or cultural program in the library	107/513
x_{22}	Participation in an educational or cultural program outside the library	44/576
x_{23}	Spending time in the library for personal relaxation	254/366
x_{24}	Spending time in the library for studying your own documents	93/527
x_{25}	Meeting with friends	72/548
x_{26}	Use of electrical plug sockets for your own computer	56/564
x_{27}	Use of library computers to access the Internet	102/518
x_{28}	Use of Wi-Fi connection in the library	104/516
x_{29}	Use of technical equipment of the library	55/565
x_{30}	Use of library for Opencard related services	104/516

Appendix 1 includes the current typology of services of the Municipal Library of Prague. It was prepared by professional library staff. They reflect the experiences of the regular operation and statutory reporting they perform. The library typology has yet to be verified, and it is not clear whether it is generally applicable to other libraries. By an appropriate empirical sample and application of statistical methods it is possible to confirm or deny the appropriateness of the classification of certain services into groups.

The typology corresponds to the services included in Table 1 as follows:

- Charge-out loan services: x_1, x_2
- In-house loan services : x_3, x_4
- Copying and printing : x_5, x_6, x_7, x_8
- Digital services : $x_9, x_{10}, x_{11}, x_{12}, x_{14}, x_{15}$
- Information, research : $x_{13}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}$
- Cultural and educational events: x_{21}, x_{22}
- Spending time in the library: x_{23}, x_{24}, x_{25}
- Technical and other services: $x_{26}, x_{27}, x_{28}, x_{29}, x_{30}$

Fig. 1 shows the frequency of use of at least one of the services of the proposed typology. From the total of 620 readers the most used services are charge-out loan services (about 97%). More than half of the readers use digital services and Information and research. Roughly half of them used the in-house loan services and the service spending time in the library. About 40% of the readers use at least some of the technical services, while the least used services were copying and printing, and cultural and educational events.

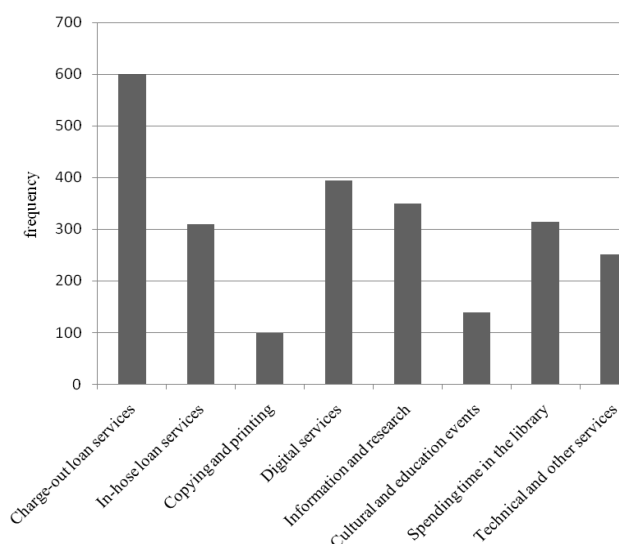


Fig. 1: Absolute frequency of use of individual types of services

The socio-economic and demographic characteristics of the users are included in Table 2. About 80% of the users were women. The sample of users was dominated

by persons aged 40 years and over (17% aged 40-49 and 30% aged 50 and over). Regarding education, about 37% had graduated from secondary school and 47% had a university degree. Most users were economically active (55%). In addition, students are represented (21%) and pensioners (16%), housewives (7%) and the unemployed (less than 1%).

Table 2: Socio-economic and demographic characteristics of the users

Abbrev.	Characteristic	Frequency
x_{31}	Gender	Male: 126 Female: 494
x_{32}	Age	15-17: 17 18-19: 28 20-24: 82 25-29: 73 30-34: 56 35-39: 72 40-49: 104 50 and over: 188
x_{33}	Education	Basic or apprenticeship: 46 Secondary: 230 Post-secondary: 51 University: 293
x_{34}	Socio-economic status	Student: 133 Housewife, maternity leave/parental leave: 42 Pensioner: 99 Unemployed: 2 Economically active: 344

The frequency of visits by users and their loan activity in the library are presented in Table 3. District branches of the library are used more often. Approximately 60% of readers use only or mostly district branches of the library. In contrast, less than 7% of readers only use the central library. The frequency of visits is dominated by the readers who visit the library about once a month (68%). Most users loan 3-6 books per month (58%).

Table 3: Frequency of visits of the users

Abbrev.	Characteristic	Frequency
x_{35}	Library branch	Only district: 155 Mostly district: 219 Both the same: 115 Mostly central: 88 Only central: 43
x_{36}	Visit frequency	Less than once per month: 96 About once per month: 422 One per week and more: 102
x_{37}	No. of loaned items per month	0: 36 1-2: 131 3-4: 193 5-6: 166 7-10: 66 11 and more: 28

4.1 Attribute extraction methods

When large multivariate datasets are analyzed, it is often desirable to reduce their dimensionality. Principal component analysis (PCA) is one of the techniques for accomplishing this. For the same purpose, independent component analysis, neural networks, and other methods are also available (see e.g. [15]). PCA replaces the original q variables by a smaller number f of derived variables, the principal components, which are linear combinations of the original variables. However, this can be impossible in some cases. Since PCA has already been described by other authors, only a brief notion of PCA is given here.

Let \mathbf{X} represent a data matrix $n \times q$, where n is the number of observations and m is the number of variables. Principal component analysis represents an optimal factorization of \mathbf{X} into two matrices, \mathbf{T} (scores matrix, $n \times f$, f is the number of components, $f < q$) and \mathbf{P} (loadings matrix, $q \times f$), plus a matrix of residuals \mathbf{E} ($n \times q$). The PCA model can be defined as follows:

$$\mathbf{X} = \mathbf{T} \times \mathbf{P}^T + \mathbf{E}. \quad (1)$$

The principle of this method lies in the fact that a variable with a higher variation explains a higher proportion of the variation in the dependent variable compared to a variable with a lesser variation. So, the original set of variables x_1, x_2, \dots, x_q is transformed into the set of components t_1, t_2, \dots, t_f . The components t_1, t_2, \dots, t_f are uncorrelated and represent most of the original variation. The condition for an optimal solution is that the Euclidean norm of the residual matrix $\|\mathbf{E}\|$ must be minimized for the given number f of components t_1, t_2, \dots, t_f . In order to reach the optimal solution, the columns of the matrix \mathbf{P} are the eigenvectors corresponding to the f largest eigenvalues of the covariance matrix \mathbf{X} .

4.2 Clustering methods

For the identification of services of the public library, K-means and hierarchical algorithms were used. These methods belong to the group of unsupervised learning methods, and also clustering. Clustering methods are used in data mining to identify groups.

The K-means method is a non-hierarchical clustering method. These methods are preferable for discovering knowledge in databases because they allow more efficient processing and interpretation of large data sets. In contrast, hierarchical clustering methods are used to create a tree structure of clusters, called a dendrogram. Clustering results are affected by outlying values and unwanted previous combinations of samples remain in the analysis.

In non-hierarchical algorithms (such as K-mode or K-means algorithms) the samples are a predetermined number of clusters. In the case of K-means algorithms initial cluster centers are set first and then the samples, located within a given distance from the center of the cluster, are assigned to the cluster. The key step here is the initial setup of the centers of the clusters (here carried out using a hierarchical single linkage algorithm), which

enables the efficient functioning of the K-means algorithm and reduces the possibility of the error (utility function being stuck at the local minimum. The aim of K-means algorithms is to minimize the utility function J , which has the form:

$$J = \sum_{r=1}^m \sum_{i=1}^n \|\mathbf{p}_i^r - \mathbf{c}_r\|^2 \quad (2)$$

where \mathbf{p}_i^r is the sample i belonging to the cluster r and \mathbf{c}_r is the center of the cluster r .

In the hierarchical clustering the Euclidean metric was chosen as a measure of dissimilarity, which can be expressed as follows:

$$\rho_E(x_i, x_j) = \sqrt{\sum_{k=1}^q (x_{i,k} - x_{j,k})^2} \quad (3)$$

where $x_{i,k}$ and $x_{j,k}$ are the values of the sample i or the sample j , for the attribute k , $k = 1, 2, \dots, q$. This distance can be also used for binary symmetric data, providing highly correlated values with other metrics such as the coefficient of agreement. Agglomerative algorithms are also used for hierarchical clustering, where the two closest samples are clustered together in each step. Specifically, we will apply the method of nearest neighbor (single linkage) and Ward's method. In the case of nearest neighbor, the clustering distance is determined by the distance of the two nearest samples from various different clusters. In Ward's method, clusters are merged with the minimum increase in the intra-group sum of squares of deviations of individual values from the cluster average.

5 RESULTS AND THEIR ANALYSIS

Fig. 2 shows the determined factors t_1, t_2, \dots, t_f and the proportion of variance in the data explained by these factors. The first component t_1 explains 13.79%. The first 10 components t_1, t_2, \dots, t_{10} have a number greater than 1 and together explain 54.45% of the variance of the original data.

Table 4 only shows components t_1, t_2, \dots, t_{10} with a number greater than 1. Attributes that have a weight greater than 0.2 or less than -0.2 are marked in bold i.e. they are strongly correlated (positively or negatively) with a particular component.

Individual components can be described verbally based on the contributions of input attributes. The first component represents all services except basic services x_1 , i.e. loan of book, magazines, CD retrieved without the assistance of a librarian. It should be noted that no one in the studied sample used the service x_{17} , preparation of written research, therefore it was not used in the analysis of the main components. This result shows that there is a large proportion of users who use service x_1 , but only in combination with certain other services.

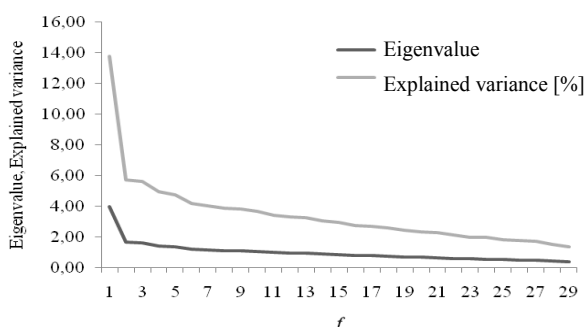


Fig. 2: Numbers and proportion of explained variance for the changing number of factors f

Other primary services (as well as secondary and tertiary services) can be replaced by component t_1 , where the greatest weight is given to electronic services x_{14} , x_{26} , x_{27} and x_{28} . Component t_2 can be called services without the assistance of a librarian (the largest positive contribution to this component is from services x_1 , x_3 , x_5 , x_7 , x_{24} and x_{28}). In contrast, these components have a negative correlation with the component with the assistance of a librarian (x_2 , x_{16} and x_{18}). Component t_3 can be described as working with documents with the assistance of a librarian (x_6 , x_8 and x_{10}). The component negatively influences the service search on the computer (x_{14} and x_{15}). Component t_4 represents the service spending time in the library without the assistance of a librarian (x_3 and x_{23}), component t_5 work with printed documents (x_7 and x_{13}) and component t_6 services with the study of documents (x_4 , x_5 , x_9 and x_{24}). Loan services based on searching in a paper catalogue are represented by component t_7 (x_1 , x_3 and x_{13}). Component t_8 combines charge-out loan services with reading electronic documents (x_1 and x_{10}), component t_9 can be described as spending time in the library with the assistance of a librarian (x_4 , x_{11} and x_{23}) and component t_{10} as loan services based on searching in the electronic catalogue combined with access to the Internet.

Table 4: Contributions of the attribute to the individual components

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
x_1	-0.02	0.22	-0.20	0.18	0.04	-0.51	0.24	0.29	0.00	0.20
x_2	0.41	-0.48	-0.25	-0.16	0.04	0.15	-0.22	0.08	-0.15	-0.14
x_3	0.40	0.28	0.00	0.40	0.01	0.05	0.31	0.10	0.18	-0.10
x_4	0.44	-0.23	0.12	-0.06	-0.11	0.26	0.12	0.12	0.29	-0.11
x_5	0.41	0.21	-0.02	0.01	0.30	0.28	-0.28	0.25	-0.05	-0.01
x_6	0.27	-0.12	0.42	-0.15	0.13	-0.05	-0.06	0.11	0.23	0.28
x_7	0.24	0.32	0.11	0.18	0.52	0.15	0.00	-0.13	0.07	0.17
x_8	0.31	-0.04	0.56	-0.24	-0.08	0.02	0.03	0.26	0.01	0.11
x_9	0.26	0.01	0.07	0.15	0.12	0.33	-0.34	-0.12	-0.35	-0.03
x_{10}	0.20	0.07	0.41	-0.42	0.05	-0.18	0.03	0.32	-0.15	-0.06
x_{11}	0.14	0.19	-0.28	-0.31	0.25	-0.03	-0.08	-0.07	0.49	-0.01

x_{12}	0.17	0.14	-0.07	-0.14	0.37	-0.18	-0.16	0.07	0.15	-0.64
x_{13}	0.22	0.00	0.23	0.16	0.41	-0.05	0.33	0.11	-0.17	-0.31
x_{14}	0.53	-0.01	-0.47	-0.05	0.15	-0.03	0.13	0.14	-0.13	0.18
x_{15}	0.42	-0.09	-0.50	-0.21	0.17	-0.24	-0.04	0.06	-0.06	0.22
x_{16}	0.41	-0.56	-0.16	0.07	0.08	-0.08	0.01	0.18	0.06	0.12
x_{17}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
x_{18}	0.38	-0.50	0.10	0.13	0.07	-0.14	0.11	-0.22	0.17	-0.10
x_{19}	0.42	-0.39	0.14	0.10	0.05	-0.14	0.12	-0.38	-0.03	-0.04
x_{20}	0.18	0.11	0.09	-0.47	-0.12	-0.34	0.05	-0.26	0.02	-0.16
x_{21}	0.23	-0.02	0.04	0.27	-0.11	-0.12	-0.44	0.25	0.18	0.13
x_{22}	0.18	0.18	0.12	0.06	-0.13	-0.40	-0.41	0.05	-0.22	-0.05
x_{23}	0.36	0.12	0.04	0.47	-0.22	-0.09	-0.03	0.01	0.29	-0.14
x_{24}	0.52	0.27	-0.01	-0.17	-0.12	0.29	0.15	0.03	0.04	0.11
x_{25}	0.38	0.02	-0.09	0.20	-0.41	-0.08	-0.19	0.16	0.03	-0.25
x_{26}	0.57	0.21	-0.11	-0.16	-0.38	0.11	0.11	-0.14	-0.10	-0.07
x_{27}	0.58	0.05	0.27	0.08	-0.08	-0.06	0.04	-0.09	-0.01	0.23
x_{28}	0.59	0.29	-0.18	-0.17	-0.25	0.11	0.11	-0.13	-0.08	-0.06
x_{29}	0.46	0.09	0.10	0.20	0.15	-0.19	0.07	-0.19	-0.43	0.04
x_{30}	0.20	0.17	0.11	-0.01	0.11	-0.14	-0.33	-0.48	0.22	0.18

The number of clusters for the K-means algorithm was determined by the shape of the utility function J for different numbers of clusters m . If the number of clusters was $m > 10$ then here is no further significant decrease in the utility function J , so the number of clusters was set to $m=10$. The results of the clustering with the K-means algorithm are presented in Table 5. The primary distance of loan services are in clusters c_7 (search without the assistance of a librarian) and c_8 (search with assistance of a librarian). Cluster c_9 represents the service spending time in the library. Cluster c_3 represents in-house services requiring the assistance of a librarian. Copying, digital and reference services are located in a cluster c_2 . Cluster c_{10} represents searching in the electronic catalogue and cluster c_6 is searching with the assistance of a librarian. Cluster c_4 is spending time in the library with your own computer and cluster c_1 represents the services associated with Opencard.

Table 5: Grouping of attributes into clusters – K-means algorithm

Cluster	Attributes
c_1	x_{30}
c_2	$x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{17}, x_{20}, x_{22}, x_{29}$
c_3	x_4, x_{19}, x_{27}
c_4	$x_{24}, x_{25}, x_{26}, x_{28}$
c_5	x_{21}
c_6	x_{16}, x_{18}
c_7	x_1

c_8	x_2
c_9	x_3, x_{23}
c_{10}	x_{14}, x_{15}

In the case of hierarchical clustering the following methods were used: single linkage and Ward's method.

Fig. 3 shows the dendrogram of the algorithm of simple connections where service x_1 creates a separate cluster. Another cluster is formed by the service searching in the electronic catalogue (x_{14} and x_{15}). In the next step we can find a cluster for the services spending time in the library (x_3 and x_{23}), etc. From the dendrogram it is also clear that the first services to cluster were those with the lowest number (x_{10}, x_{17}, x_{20} , etc.) which are similar from the point of view of their low frequency of use. Table 6 shows the services belonging to the clusters that were derived from the dendrogram in Fig. 3. Here the number of clusters depends on the choice of the user and can be determined on the basis of changes in the distance of the connections. A large distance separates the number of clusters 9 and 10, so the number 9 was selected. The differences compared to the results achieved by the K-means algorithm are as follows. Charge-out loans are in clusters c_1 and c_2 . Distance loans are in cluster c_6 . Spending time at the library services are in cluster c_7 . Information and retrieval services are in clusters c_4, c_5, c_8 and c_9 . Finally, the remaining services are concentrated in cluster c_3 , i.e. copying and printing, most of the digital services, research services, other spending time in the library services, technical services, and cultural and educational events.

Table 6: Grouping of attributes into clusters – single linkage

Cluster	Attributes
c_1	x_1
c_2	x_2
c_3	$x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{17}, x_{19}, x_{20}, x_{21}, x_{22}, x_{24}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}, x_{30}$
c_4	x_{18}
c_5	x_{16}
c_6	x_3
c_7	x_{23}
c_8	x_{14}
c_9	x_{15}

Compared to the previous methods the Ward's method provides a key contribution to the variance rather than the distance of the samples. This is reflected in the different shape of the dendrogram (Fig. 4) where e.g. the primary service charge-out loans (x_1 and x_2) are clustered together with spending time in the library (x_3 and x_{23}) and retrieval services (in particular searching in the electronic catalogue). The rest consists of other services. In the next step the services can be divided into charge-out loan services without the assistance of a librarian together with spending time in the library and retrieval services together with charge-out loans with the assistance of a librarian. Similarly, other services can be divided into two major clusters, i.e. cultural and educational services and electronic services in one cluster and copying and reference services in the other cluster.

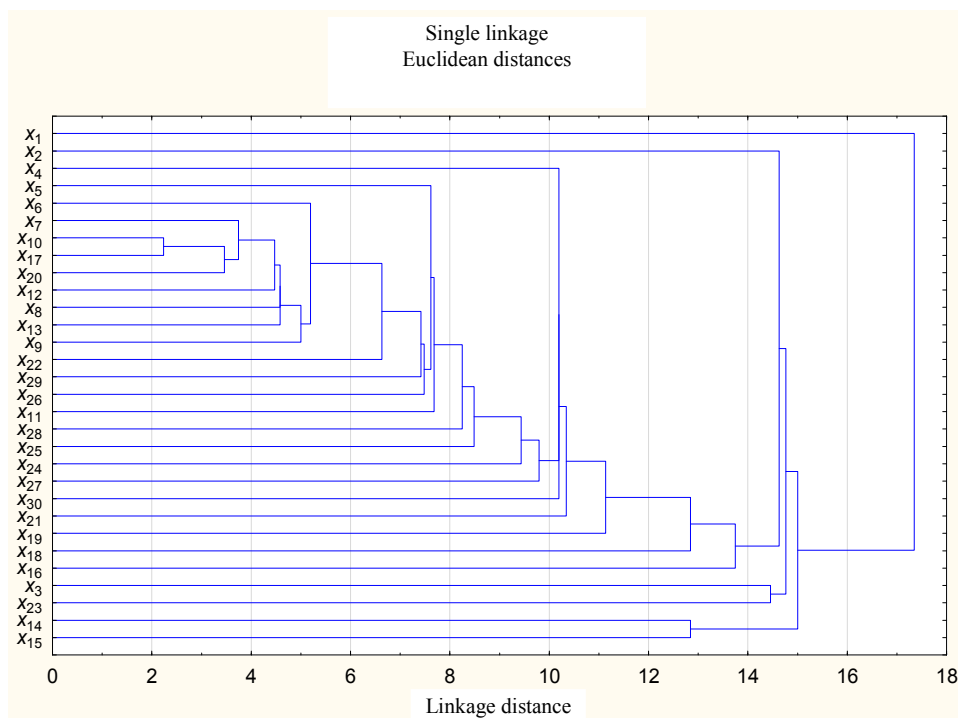


Fig. 3: Dendrogram for single linkage

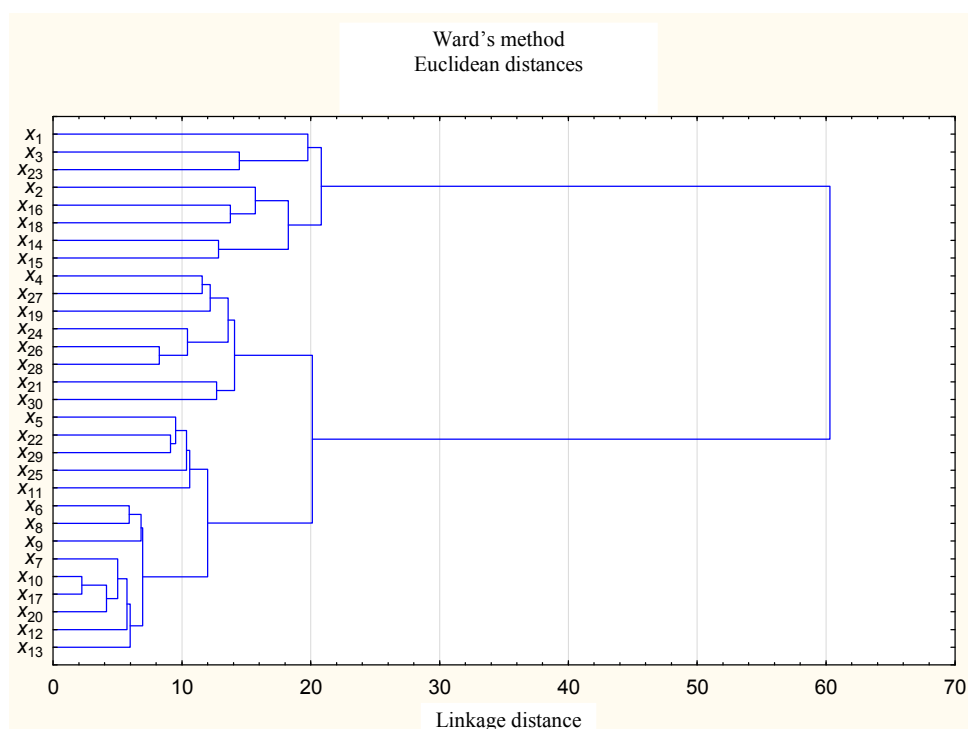


Fig. 4: Dendrogram for Ward's method

The results of the Ward's method suggest that services can be divided into two large clusters. Grouping of the services in these clusters is shown in Table 7. Cluster c_1 includes both charge-out and in-house loan services retrieved without the assistance of a librarian, information services and spending time in the library. In cluster c_2 are copying and printing services, digital services, cultural and educational events, other spending time in the library and technical services. Cluster c_1 can be described as the basic services that should be provided by each public library (primary services) and cluster c_2 secondary services and others. Ward's algorithm therefore confirmed the existence of primary and secondary services at the library, whereas the algorithms for simple and total connection only divided the services based on frequency of use. These algorithms classified the services that are used most frequently into individual clusters. Other lesser used services were grouped into one large cluster of services.

Table 7: Grouping of attributes into clusters – Ward's method

Cluster	Attributes
c_1	$x_1, x_2, x_3, x_{14}, x_{15}, x_{16}, x_{18}, x_{23}$
c_2	$x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{17}, x_{19}, x_{20}, x_{21}, x_{22}, x_{24}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}, x_{30}$

6 CONCLUSION

The results obtained using the statistical methods indicated that the primary, secondary and tertiary library

services are provided in combination and complement each other depending on the needs of the users. For future research methodology it was confirmed that Ward's method provides an appropriate mechanism for the connection of attributes (types of services) into the appropriate primary and secondary groups as the most common grouping of library services.

Monitoring the use of services without the assistance of a librarian (loans, spending time in the library, electronic services) on the one hand and the same services with the assistance of a librarian on the other appears to be of key importance to the management of the library. Especially for off-site and full-time loan services is a key division according to whether they are searched, with or without the assistance of a librarian.

The K-means method verified the clustering of library services and thus verifies the methodology from the Municipal Library of Prague. The result is a proposal of grouping of services as follows:

- charge-out loan services with the assistance of a librarian (x_1),
- charge-out loan services without the assistance of a librarian (x_2),
- copying and printing (x_5, x_6, x_7, x_8),
- relaxation, rest, incl. rest through reading your own books (x_3, x_{23}),
- relaxation, rest, incl. rest through reading library books with the assistance of a librarian and access to the internet via a wired connection (x_4, x_{19}, x_{27}),
- cultural and educational events (x_{21}, x_{22}),

- study, meetings, computer operation, incl. internet through Wi-Fi (x_{24-26}, x_{28}),
- research, searching for the availability of books (x_{16}, x_{18}),
- catalogue search on computers or the website (x_{14-15}).

By further grouping these groups it is possible to validate the basic intentions of the individually predefined groups of service as follows:

- loan services,
- rest and relaxation connected with time spent in the library,
- information and research,
- cultural and educational events,
- technical and digital services.

In these groups it would of great benefit to analytically monitor performance and cost and to perform an economic and output analysis. For verification it is necessary to recommend additional research of the services provided by libraries through the use of other statistical or computational intelligence methods such as neural networks with unsupervised learning [16,17] or independent component analysis [15]. Non-linear relations among library services can be detected using these methods.

ACKNOWLEDGEMENTS

This paper was prepared as part of the research project No. 3052/2012 SOKL entitled “Methodology for measuring the value of library services” funded by the Ministry of Culture of the Czech Republic.

REFERENCES

- [1] S. Nicholson, The Bibliomining Process: Data Warehousing and Data Mining for Library Decision-Making, *Information Technology and Libraries*, Vol.22, No.4, 2003, pp.146-151.
- [2] S. Nicholson, The Basis for Bibliomining: Frameworks for Bringing together Usage-based Data Mining and Bibliometrics through Data Warehousing in Digital Library Services, *Information Processing and Management*, Vol.42, 2006, pp.785-804.
- [3] A. Geyer-Schulz, A. Neumann, A. Thede, An Architecture for Behavior-Based Library Recommender Systems, *Information Technology and Libraries*, Vol.22, No.4, 2003, pp.165-174.
- [4] K. Guenther, Applying Data Mining Principles to Library Data Collection, *Computers in Libraries*, Vol.20, No.4, 2000, pp.60-63.
- [5] S.C. Kao, H.C. Hang, C.H. Lin, Decision Support for the Academic Library Acquisition Budget Allocation via Circulation Database Mining, *Information Processing and Management: an International Journal*, Vol.39, No.1, 2003, pp.133-147.
- [6] D. Pierrakos, G. Paliouras, C. Papatheodorou, C.D. Spyropoulos, Web Usage Mining as a Tool for Personalization: A Survey, *User Modeling and User-Adapted Interaction Journal*, Vol.13, No.4, 2003, pp.311-372.
- [7] C. Papatheodorou, S. Kapidakis, M. Sfakakis, A. Vassiliou, Mining User Communities in Digital Libraries, *Information Technology and Libraries*, Vol.22, No.4, 2003, pp.152-157.
- [8] E. Garfield, A.I. Pudovkin, V.S. Istomin, Mapping the Output of Topical Searches in the Web of Knowledge and the Case of Watson-Crick, *Information Technology and Libraries*, Vol.22, No.4, 2003, pp.183-187.
- [9] R. Kostoff, J. del Rio, J. Humenik, E. Garcia, A. Ramirez, Citation Mining: Integrating Text Mining and Bibliometrics for Research User Profiling, *Journal of the American Society for Information Science and Technology*, Vol.52, No.13, 2001, pp.1148-1156.
- [10] Ohio, *Value for Money: Southwestern Ohio's Return from Investment in Public Library*, Driscoll & Fleeter, Columbus, OH, 2006.
- [11] British Library, *Measuring our Value*, British Library Annual Report 2003/2004, 2004, available online at: <http://www.bl.uk/pdf/measuring.pdf>.
- [12] T. Lynch, *A Study of Taxpayer Return on Investment (ROI) in Florida Public Libraries: Part III - REMI Details*. Tallahassee, FL: State Library and Archives of Florida, 2004.
- [13] J.R. Matthews, *Measuring for Results – The Dimensions of Public Library Effectiveness*, Libraries Unlimited, London, 2004.
- [14] J. Stejskal, P. Kotatkova Stranska, K. Matatkova, P. Hajek, Public Services Value Determining – Case of Public Libraries. In *Proceedings from 4th WSEAS World Multiconference on Applied Economics, Business and Development (AEBD 12)*, Porto, 2012, pp.140-145.
- [15] I. Guyon, S. Gunn, M. Nikravesh, L.A. Zadeh, *Feature Extraction – Foundations and Applications*, Springer Verlag, Berlin, Heidelberg, New York, 2006.
- [16] V. Olej, P. Hajek, Modelling Municipal Rating by Unsupervised Methods. *WSEAS Transactions on Systems*, Vol.5, No.7, 2006, pp.1679-1786.
- [17] V. Hajkova, P. Hajek, Typology of Regional Innovation Systems in Europe - A Neural Network Approach. *International Journal of Mathematical Models and Methods in Applied Sciences*, Vol.5, No.3, 2011, pp.463-471.

Petr Hajek is an associate professor with the Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Czech Republic. He deals with the modeling of economic processes (especially in the field of public economics and public finance), advanced statistical methods, and methods of computational intelligence (soft-computing). (e-mail: petr.hajek@upce.cz).

Now Assoc. Prof. Hajek is co-investigator of the grant financed from Ministry of Culture of Czech Republic under the name “Methodology for measuring the value of library services” during 2011 - 2012 .

Jan Stejskal is an associate professor in public and regional economics. Studied in University of Pardubice, Faculty of Economics and Administration in Pardubice, Czech Republic. His domain is connection of the public economy in the regional scope and view. In the papers he describes and analyzes specially the regional policy, tool of the local and regional economic development, public policies after care of the investors, systems of support in national economy in EU countries.

Dr. Stejskal is senior lecturer in University of Pardubice, Faculty of Economics and Administration. He teaches subjects: Economy of Public Sector, Management of Public Sector, Public Economics for bachelor and master studies. He is author/co-author of many papers and also monographs, for example Regional policy and its tools (Prague, Czech Republic: Portal, 2010 – in Czech), Economic Development of Regions in the Context of Sustainability (Saarbrücken, Germany: Lambert Academic Publishing, 2010), The Industrial Clusters and Their Birth in Regions (Prague, Czech Republic: Linde, 2011 – in Czech). Now Dr. Stejskal is investigator of the grant financed from Ministry of Culture of Czech Republic under the name “Methodology for measuring the value of library services”.

APPENDIX 1

Result of the typology of services from the Municipal Library of Prague

- *Charge out loan services - home lending*
 - Loan of books, magazines, CD retrieved without the assistance of a librarian;
 - Loan of books, magazines, CDs from stock, retrieved with the assistance of a librarian;
- *In-house loan services - in the library*
 - Reading books, newspapers or magazines, etc. studying documents etc. in the library - retrieved without the assistance of a librarian;
 - Reading of books, newspapers or magazines, etc. study of documents etc. in the library - retrieved with the assistance of a librarian;
- *Copying and printing*
 - Copying or printing of documents (with or without the assistance of a librarian);
- *Digital services - in the library building*
 - Catalogue search in the library - Koniáš;
 - Use of specialized electronic databases (e.g. Proquest, EBSCO, Anopress, Newton) on computers inside the library or in the library building (except for use of the library catalogue Konias);
 - Reading e-books and other electronic documents on computers in the library or in the library building;
 - Use of computers in the library for Internet access or for office applications, image editing, etc. (not for work with catalogues and databases);
- *Digital services - outside the library*
 - Catalogue search on the MLP website;
 - Downloading e-books from the MLP website;
 - Downloading/viewing articles and other electronic documents from professional databases (e.g. Proquest, EBSCO) via the website (except for use of the library catalogue Koniáš);
- *Information, retrieval*
 - Catalogue search in the card or paper catalogues in the library;
 - Requesting assistance from a librarian in finding information in the library or on the Internet, fact finding (except for questions on the library services and the books in the library);
 - E-mail or telephone requests for assistance in finding information, fact finding (except for questions on the library services and the books in the library);
- *Cultural and educational events*
 - Participation in educational or cultural programs in the library or programs held by the library outside the library (e.g., World of Books, Bambiriáda etc.);
- *Spending time in the library*
 - Spending time in the library for personal relaxation, study of documents other than library materials or meeting and talking with friends, etc. (i.e., time spent in the library more than the time required for processing the loan and the other above mentioned library services);
- *Technical and other services*
 - Use of electrical plug sockets for your own computer, tablet or similar device;
 - Use of Wi-Fi connection in the library;
 - Use of other technical equipment of the library - players, digital piano, game consoles, etc.;
 - Use of MLP for Opencard related services.